SUBSCRIBERS'

APR 10 1921

Vol. IV, No. 1 PSYCHOLOGICAL REVIEW PUBLICATIONS FEBRUARY, 1921

Journal of Experimental Psychology

EDITED BY

JOHN B. WATSON, 344 Madison Ave., New York City
HOWARD C. WARREN, Princeton University (*Review*)
JAMES R. ANGELL, 525 Fifth Ave., New York City (*Monographs*)
WALTER S. HUNTER, Gen. Hosp. for Insane (*Bulletin*) and
HERBERT S. LANGFELD, University of Illinois (*Index*)

ASSISTANT EDITORS

E. A. [illegible], University of Chicago; RAYMOND DODGE, Wesleyan University; KNIGHT DUNLAP, [illegible] University; H. S. LANGFELD, Harvard University; E. L. [illegible]

CONTENTS

The [illegible] Young Trees and Fairchild. HAROLD E. BURTT, 1.

An [illegible] Ability, F. A. C. PERRIN, 24.

[illegible] and Sound. P. L. WELLS, C. M. KELLEY, [illegible]

[illegible] Error: SAMUEL W. FERNBERGER, 63

REVIEW COMPANY
[illegible], PA., AND

[illegible] Association, Inc., proprietor

INDEX FOR 1921

February.

The Inspiration-Expiration Ratio During Truth and Falsehood: HAROLD E. BURTT, 1.
An Experimental Study of Motor Ability: F. A. C. PERRIN, 24.
Comparative Simple Reactions to Light and Sound: F. L. WELLS, C. M. KELLEY, GARDNER MURPHY, 57.
An Experimental Study of the "Stimulus Error": SAMUEL W. FERNBERGER, 63.

April.

The Pursuitmeter: WALTER R. MILES, 77.
Further Technique for Inspiration-Expiration Ratios: HAROLD E. BURTT, 106.
Complex Signs in Diagnostic Free Association: C. L. HULL and L. S. LUGOFF, 111.
Effects Simulating Fatigue in Simple Reactions: F. L. WELLS, C. M. KELLEY, GARDNER MURPHY, 137.
Subjective Perceptions: KARL M. DALLENBACH, 143.

June.

A Mirror Recorder for Photographing the Compensatory Movements of Closed Eyes: RAYMOND DODGE, 165.
A Method for Measuring Retinal Sensitivity: PERCY W. COBB and MILDRED W. LORING, 175.
Improvement in Brightness Discrimination and its Bearing on a Behavioristic Interpretation of Perception: EDWARD S. JONES, 198.
Physical Attractiveness and Repulsiveness: F. A. C. PERRIN, 203.
Apparatus for the Study of Visual After-Images: DONALD A. LAIRD, 218.
The Modification of Intensity of Sensation by Attention: S. M. NEWHALL, 222.
An Improvement in Voice Keys: KNIGHT DUNLAP, 244.

August.

The Latent Time of Compensatory Eye-Movements: RAYMOND DODGE, 247.
An Experimental Study of Associative Inhibition: LINUS W. KLINE, 270.
The Effects of Practice upon the Scores and Predictive Value of the Alpha Intelligence Examination: FLORENCE RICHARDSON and EDWARD S. ROBINSON, 300.
The Taste Sensitivity of an Anosmic Subject: A. R. GILLILAND, 318.

October.

The Relative Efficiencies of Distributed and Concentrated Study in Memorizing: EDWARD S. ROBINSON, 327.
The Colors Produced by Equilibrium Photopic Adaptation: LEONARD THOMPSON TROLAND, 344.
On Attention and Simple Reaction: F. L. WELLS, C. M. KELLEY and GARDNER MURPHY, 391.

404413

December.

The Influence of Visual Guidance in Maze Learning: HARVEY CARR, 399.
Smaller vs. Larger Units in Learning the Maze: J. W. BARTON, 418.
Improved Forms of Steadiness Tester and Tapping Plate: KNIGHT DUNLAP, 430.
Eliminating the Pitfalls in Solving Correlation—A Printed Correlation Form: HERBERT A. TOOPS, 434.
The Role of Synæsthesia in Learning: RAYMOND HOLDER WHEELER and THOMAS D. CUTSFORTH, 448.
Local Signature and Sensational Extensity: W. C. RUEDIGER, 469.

Journal of
Experimental Psychology

VOL. IV, NO. 1. FEBRUARY, 1921

THE INSPIRATION–EXPIRATION RATIO DURING TRUTH AND FALSEHOOD

BY HAROLD E. BURTT

Ohio State University

1. INTRODUCTION

Three variables have been studied as diagnostic of truth and falsehood: the association reaction, systolic blood pressure, and the inspiration-expiration ratio. Benussi[1] obtained rather clear results with the last of these methods. The writer in collaboration with Marston and Troland made a brief study of all three methods with a view to military use.[2] In this case the breathing results were negative. Since that time a technique has been developed which increases the accuracy and ease of obtaining the I/E (ratio of inspiration time to expiration time). The present study utilized this improved technique to repeat in its essential features and extend Benussi's experiment.

Benussi's subjects were given small cards containing a number of digits and letters and a picture. They were questioned as to the nature, number and arrangement of the symbols and were asked to describe the picture and to read the symbols in a specified order. In some instances they did this truthfully but with other cards (marked with a star) they lied upon every point. A number of spectators were present to enhance the subjects' emotional state. The

[1] Benussi, V., 'Die Atmungsymptome der Luge,' *Archiv f. die gesamte Psychologie*, **31**, 1914, 244–273.

[2] Cf. Shepard, J. F., 'Review of Affective Phenomena,' *Psych. Bull.*, **16**, 1919, 244.

breathing was recorded on a kymograph and several breaths immediately preceding and following each reply were measured to compute the I/E.

Benussi found that the average of 3 to 5 ratios preceding a false answer was less than the average of 3 to 5 ratios following that answer. With truthful replies the reverse was the case, —the average I/E before the answer was greater than the average I/E after the answer. The difference between the average ratios before and after a lie varied with the ability of the subject to dissimulate (as judged by the experimenter) and correlated also with the difficulty of lying in specific instances as indicated by the introspection. The results were negative if the audience knew whether the subject was going to lie or not and if the subject was aware of this. Benussi found that when the subject tried to voluntarily control the breathing, even following a metronome, the I/E was still diagnostic.

2. Method

The subject in the present experiment wore a pneumograph of the usual type,—a large rubber tube with a spiral spring inside. This connected pneumatically with a tambour which closed an electric circuit during inspiration and opened it during expiration.[1] In the earlier part of the experiment this operated a double relay which closed circuits through two signal magnets,—one during inspiration and one during expiration. These two circuits were interrupted by a vibrator 5 times per second. The magnets recorded in a spiral on a kymograph about 12 feet distant from the subject and the I/E was obtained by counting vibrations on the two lines. With a telegraph key in parallel with the relay circuits the experimenter could mark on the record in simple code the number of the question and other data necessary for interpretation. In the later part of the experiment the relay circuits operated a device which indicated the I/E at the close of each expiration[2] so that it could be recorded immedi-

[1] Burtt, H. E., 'A Pneumograph for Inspiration-Expiration Ratios,' *Psychol. Bull.*, 15, 1918, 325–328.

[2] Burtt, H. E., 'Further Technique for Inspiration Expiration Ratios,' Jour. of Exper. Psych. (to appear shortly).

ately. It consisted essentially of two light cars moving at right angles from a common point, one during inspiration and the other during expiration. A pointer was pivoted on the former and rested against a stop on the latter so that at the end of a breath the I/E was a function of the angle covered by the pointer and could be read from a scale. The subject also wore a sphygmomanometer of the portable type and a bracelet stethoscope. Systolic blood pressure readings were taken at intervals.

The experiment consisted of four series. The first employed material quite similar to Benussi's. A set of 3 x 5 inch cards was prepared each containing capital letters or digits or both. There were from 4 to 8 symbols on a card and they were arranged in one of 6 simple geometrical shapes,—triangle, circle, etc. The number, nature and shape of the symbols was determined by chance and 18 cards of the sort were prepared. The subject after 30 seconds study was questioned on four points: the geometrical shape, the kind of material (letters or numbers or both), the number of symbols, and an actual reading of the symbols in some prescribed fashion. The order of these questions was varied. The subject had before him a reference card on which were the six geometrical shapes, a statement of the fact that there were always from 4 to 8 symbols and the four points above mentioned on which he was to be questioned. He was informed that if, when lying, he confined himself to the limits indicated the experimenter could not tell from the answers whether he was lying or telling the truth about a card selected at random from the group. In an hour's experiment eight cards were generally used. Four of these the subject described truthfully and with four he lied on every point. The order was determined by chance and unknown to the experimenter.

The above points were explained to the subject and he was told to try to deceive the experimenter. The latter after each card stated whether he considered the performance was truth or lie. This was primarily to create greater interest and perhaps emotional reaction. One or more spectators were sometimes present for the same purpose. They likewise passed judgment after each card.

In order to obtain breathing records before and after the answers the subject was instructed to remain silent when asked a question, to think about the answer until the signal 'now' after which he could reply as soon as he wished, and to then keep silent until the next reply. The experimenter waited about 10 seconds,—long enough for 3 to 5 breaths,—after asking a question before saying 'now' and waited a corresponding length of time before giving the next question. The subject was told that this procedure was to avoid his betraying himself by the length of time he took in replying to questions. Blood pressure readings were taken at the outset of the hour and after the second and fourth questions on each card.

The method of Series II. was exactly like that of Series I. just described except that the first four cards were all truth or lies and the last four the reverse. The order of the two parts of the hour's experiment was determined by chance and unknown to the experimenter. His judgment had to be deferred, of necessity, until the end of the hour.

In Series III., the subject was given two typewritten sheets of paper, face down, marked respectively T and L. He selected whichever he wished. In either instance the paper gave an account,—comprising at least 10 points on which he could be questioned,—of a crime of which he was supposed to be accused. If he chose the paper marked L the account implicated him in the crime and he had to lie concerning nearly every point in order to clear himself. The paper marked T, however, recounted the same evidence against him but provided an alibi on every point. He had merely to tell this alibi without fabricating. The following is an excerpt from an L and a T paper covering the same 'crime.'

L

You are named as corespondent by Mrs. James Wyandotte in a suit against her husband for divorce. Witnesses have established the following:

1. About Christmas time you began wearing an expensive fur although you had received no raise in salary recently and none of your friends or relatives admitted that they gave it to you.

2. About New Year's you appeared with an expensive necklace.

3. On the evening of January 10 you and Mr. Wyandotte sat in adjacent seats at the Hartman theatre and after that took lunch at the Busy Bee. You were not at home that night.

4. On the afternoon of February 24 part of a telephone conversation which you were carrying on from Hennick's was overheard. About the middle of the conversation you said: "Well what will friend wife say?" . . . etc.

T

You are named as corespondent by Mrs. James Wyandotte in a suit against her husband for divorce.

The truth is as follows:

1. About Christmas you purchased a fur with money which you made through speculation, on the strength of a tip overheard at the office where you work.

2. You purchased a necklace a little later with money from the same source.

3. You went to the Hartman Theatre, January 10, with a girl friend, Miss Helen Reilly, and it was only after you had been seated that you noticed Mr. Wyandotte's presence in the next seat. After the theatre he followed Miss Reilly and yourself to the Busy Bee where he seated himself at the same table with you and tried to flirt with you, and your friend. You spent that night with your friend, Miss Reilly.

4. The telephone conversation which was overheard at Hennick's was with this same girl friend. Your remark about "What will friend wife say?" was in regard to a party which Miss Reilly was going to hold in her rooms. 'Friend-wife' was a name which you called Miss Reilly's landlady. . . . etc.

A considerable variety of crimes were used including murder, theft, arson, forgery, fraud, treason and grounds for divorce.

The L papers were written by the experimenter but the corresponding T papers by an assistant. The experimenter did not examine the T papers and thus could not tell from the subject's replies whether he was making up his own alibi, *i.e.*, lying, or following the one written by the assistant, *i.e.*, telling the truth.

The subject was allowed 5 minutes to study the paper he selected. After each question was asked he waited for the signal 'now' before replying as in the previous series. Blood pressure was likewise measured at the outset and after every three or four questions. Two "crimes" were performed in an hour's experiment. The subject was instructed to select the L in one crime and the T in the other but the order was immaterial and unknown to the experimenter. The latter at the end of the hour made a judgment as before, to insure interest, on the basis of the subject's manner and a casual survey of the blood pressure record.

Series IV. was identical with Series III. in procedure except that a number of 'jurors' (from 3 to 26) sat facing the

subject and after each crime judged on the basis of his manner and statements whether it was T or L. This jury did not know that one performance during the hour was to be T and the other L but supposed the subject would do whatever he wished in each instance.

The questioning of the subject in Series III. and IV. was not conducted after the manner of a legal cross examination. A list of questions (unknown to the subject) was prepared before hand and asked in a matter of fact manner. Occasionally a few additional questions were interspersed. Undoubtedly an agressive rigorous examination would produce a heightened emotional state. But the aggressiveness is a variable which depends on the personality of the examiner. The writer was interested in whether this personality variable could be eliminated and still obtain positive results.

In all series for each complete breath between question and answer (there were from 3 to 5 breaths) the I/E was computed and these ratios averaged. Similarly the ratios for the 3 to 5 breaths following the answer were averaged. It seemed best, following Benussi, to average the ratios rather than to average inspirations and divide by the average of the corresponding expirations. (These methods would of course be identical if the geometric mean were to be used throughout but there is no particular justification for doing so.) It was desirable to note the ratio for each separate breath in order to study, for instance, the diagnostic accuracy of earlier and later breaths.

The average I/E for the breaths following an answer was always subtracted from the average I/E for those preceding that answer. For convenience this difference will hereafter be called D. According to Benussi, D should be positive for true answers and negative for false answers. In some cases D was computed, using only one breath before and one after the reply. These cases will be mentioned later.

The experiments were performed in the psychological laboratory of the Ohio State University, during the winter of 1919–20. Three graduate students (2 female, 1 male) served as subjects throughout. A number of other persons served as

subjects in one or two series. Of these 2 were graduate students (both male), 7 were undergraduates (2 female, 5 male) and 2 were outsiders (both female).

3. Results of Series I and II

The first two series, it will be recalled, employed cards containing letters and digits similar to Benussi's. Four questions were asked about each card and there were usually 4 L cards and 4 T cards in an hour's experiment, a total of 32 questions. Several aspects of the results of these series are presented in Table I.

Table I

Subject	Av. *D* for 16 Questions		No. of Questions *D* Correct		No. of Cards *D* Correct		B.p. Cards Correct	
	L	T	L	T	L	T	L	T
(Ser. I)								
B	−.09	.01	8/16	8/16	3/4	3/4	3/4	4/4
B	.07	.03	3/16	10/16	1/4	2/4	4/4	2/4
C	.05	−.10	5/11	3/12	0/3	0/3	3/3	3/3
E	−.11	.09	8/12	9/13	3/4	3/4	3/4	3/4
H	−.24	−.10	4/4	2/5	2/2	1/2	3/3	3/3
J	−.18	−.05	8/13	7/15	4/4	3/4	2/4	2/4
K	−.01	.00	9/14	7/13	2/4	2/4	2/4	2/4
M	.03	.16	4/10	7/13	1/4	4/4	1/4	2/4
N	.04	.06	6/16	9/15	1/4	3/4	3/4	3/4
Av. or total	−.05	.01	55/112	62/118	17/33	21/33	24/34	24/34
			49%	53%	52%	64%	71%	71%
(Ser. II)								
B	.02	.02	7/12	7/11	1/3	2/3	x	
K	.12	.12	4/13	6/11	1/4	2/4	x	
M	.07	.05	3/13	9/15	1/4	2/4	x	
N	.06	−.09	3/14	5/15	1/4	1/4	x	
Av. or total	.07	.03	17/52	27/52	4/15	7/15		
			33%	52%	27%	47%		

The data in each row are based on an hour's experiment with the subject indicated at the left. The *D*'s (average *I*/*E* before question minus average *I*/*E* after question) for the 16 questions in which the subject lied are averaged in the 2d column. Similarly the average *D* for the 16 questions in which the subject told the truth is given in the 3d column. According to the hypothesis *D* should be negative in sign for L and positive for T. Column 4 gives the number of L questions in which *D* is correct according to hypothesis, *i.e.*, negative. This figure is followed by the total number of L

questions, because in some instances the record was not complete for all 16 questions. Column 5 gives likewise the number of T questions in which *D* is correct, *i.e.*, positive, followed by the total number of T questions. These same questions are of course the basis of the averages in the preceding columns. Each card was also evaluated as a unit, *i.e.*, the 4 *D*'s of the 4 questions averaged into a single score. The number of cases in which this score correctly diagnoses an L card are given in column 6 together with the total number of L cards. Column 7 gives similar results for the T cards. The last two columns give the number of cards correctly diagnosed on the basis of systolic blood pressure. These results will be considered separately (section 6).

The upper portion of the table presents the results for Series I. in which the T and L cards occurred in a random order. Columns 2 and 3 are averaged and the remaining columns totaled. With these totals moreover the first figure in each column is taken as a per cent. of the second and these percentages given directly below (for example in the 4th column 55 is 49 per cent. of 112). The lower portion of the table presents the results for Series II. in which the 4 L cards were used in succession followed by the 4 T cards or vice versa. These results are likewise averaged or totalled. With the blood pressure readings it seemed desirable to evaluate not each card but the two groups of cards. The crosses in the last column indicate that the diagnosis as to whether the first or last four cards are L is correct.

The results for Series I. indicate a certain correspondence, although not a striking one, between the I/E and the telling of truth or falsehood. It will be recalled that in this series rather simple and unemotional stimuli were used, the questions were not of the cross-examination type and were known beforehand by the subject. Under these circumstances any correspondence between performance and breathing is encouraging. The second and third columns of Table I. show that for the majority of the subjects *D* is diagnostic and the average *D* for the L's is $-.05$ as compared with .01 for the T's. If the average of all the original *D*'s (about 115 of each

sort) is used instead of the average of the figures in the table the results are — .06 for L and .04 for T. The difference between these two averages is 4.6 times the probable error of difference and hence significant.

Considering the percentages in the following columns it is evident that in only about half of the original questions (49 per cent. on L and 53 per cent. on T) is *D* diagnostic. This apparent discrepancy with the preceding averages is probably due to the fact that many questions produce little consciousness of deception or truth whereas an occasional one does so to a considerable degree. Many of the original records show a considerable number of *D*'s varying slightly on both sides of zero, intermixed with a few which are of considerable plus or minus magnitude. The percentage of correct diagnoses is slightly greater when considering cards (column 6 and 7) rather than individual questions. It seems plausible that one of the four questions on a card might provoke a significant reaction which would alter the card average.

Two individual points should be noted. Subject *E* performed in the presence of 12 spectators and might be expected to have a heightened lying consciousness, whereas in two other series one spectator was present and in the rest no one but the experimenter. *E*'s breathing appears in the table the most diagnostic of the group. Subject *M* confessed later that instead of fabricating regarding the card before him he kept the preceding card in view and described that for his lie, thus in reality telling the truth. His average shows a positive, *i.e.*, truthful, *D* for both parts of the series.

The relative sensitivity of the first and last questions on each card was studied to determine if the lying consciousness seemed objectively more in evidence early or late during the process. The results are not included in the table because of their doubtful significance. *D* is slightly more diagnostic on the last question but this might well be due to the fact that the last question was usually more complex,—involving a reading of the symbols in some prescribed orders.

The results of series II are less convincing. The averages of columns 2 and 3 show positive values in both cases and

even a slightly higher value (.07) for L than for T (.03). The direction of this difference is the reverse of that called for by hypothesis, but a calculation using the original *D*'s shows that this difference between the averages is only 1.8 times the probable error of difference and so of very doubtful significance. The percentages in the following columns likewise show little correspondence between breathing and mendacity. There is frequent occurrence of positive *D*'s in L questions, *i.e.*, breathing indicating truth. (*Cf.* columns 4 and 6.)

Two facts might partially account for the less significant results of Series II. In the first place all the subjects had already done Series I. and perhaps had become more used to the situation and less excited by it. In the second place the four L cards were taken in succession and the lying consciousness might wear off during that process. As a matter of fact the first L card of the 4 shows a negative average *D* for 3 of the 4 subjects whereas the last L card shows a positive average *D* for all subjects.

The results of these two series seem then to indicate that even with simple unemotional material and with questions known beforehand and asked in an unexciting manner there is some evidence of correspondence between the breathing ratio and the telling of the truth or falsehood. This correspondence is generally due to a few questions in the group while the majority of them are neutral. The influence of lying upon the breathing appears to become less with habituation.

4. Results of Series III

In Series III., it will be recalled, the material consisted of imaginary crimes and in a given hour's experiment the subject lied about one crime and told the truth about another. It was thus possible to deal with absolute values or to compare a T and an L performed under practically the same conditions. The results are summarized in Table II.

The data in each row are based on one hour's experiment. The left portion of the line (columns 2 to 7) gives figures for the crime regarding which the subject lied and the right portion (columns 8 to 13) for the crime regarding which he

TABLE II

Subject	Lie						Truth						Correct Alternative			Correct b.p.
	D_1	No. Cor.	D_2	No. Cor.	D_3	No. Cor.	D_1	No. Cor.	D_2	No. Cor.	D_3	No. Cor.	D_1	D_2	D_3	
K	−.02	5/10	−.08	7/10	.01	5/10	.02	8/12	−.02	7/12	−.03	5/12	x	x	o	x
M	.04	5/11	.24	5/11	.21	4/11	−.01	6/11	.06	5/11	−.07	5/11	o	o	o	x
M	.21	2/8	.16	2/8	.29	1/8	−.13	4/12	.00	7/12	−.02	6/12	o	o	o	x
N	.01	2/10	.06	2/10	.09	1/10	.00	5/11	.06	7/11	.09	7/11	o	?	?	x
N	.02	4/6	−.02	2/6	−.02	3/6	.01	2/4	−.03	0/4	.00	0/4	o	o	x	x
N	−.06	8/12	−.02	5/12	−.03	3/12	−.03	7/12	−.09	2/12	−.11	3/12	x	o	o	o
Q	−.07	4/8	.05	2/8	−.01	3/8	.07	4/5	.06	4/5	.08	4/5	x	x	x	o
Av. or total	.02	30/65 46%	.06	25/65 38%	.08	20/65 31%	−.01	36/67 54%	.01	32/67 48%	−.01	30/67 45%	3/7	2/7	2/7	5/7

D_1: av. of 3 to 5 breaths between question and answer minus av. of 3 to 5 breaths following answer.
D_2: first breath after question minus first breath after answer.
D_3: first breath before answer minus first breath after answer.

told the truth at the same sitting. Symbols representing the crimes are not included in the table because the difficulty and emotional stimulus of the various crimes was an individual matter and impossible to evaluate. The usual procedure is to average 3 to 5 I/E ratios preceding the reply and subtract the average of 3 to 5 ratios following the reply. This is done for each question. The average of these D's on a given crime with a given subject are given in the columns marked $D1$. The immediately following columns (3 and 9) give the number of questions in which this $D1$ is correctly diagnostic (*i.e.*, negative for L and positive for T) together with the total number of questions on which ratios are available. It seemed possible that certain breaths might be more significant than others. Accordingly the I/E for the breath following an answer is subtracted from the I/E for the breath following the experimenter's question. The average of these differences for a given crime are given in the columns marked $D2$. The next columns in each instance (5 and 11) give the number of questions in which $D2$ is correctly diagnostic, and the total number of questions. Columns marked $D3$ give results similar to $D2$ using the breath just preceding the answer instead of that just following the question. Columns 14 to 16 deal with a comparison of the T and L on the same hour. Knowing that one average was made during L and one during T the assumption is that the L will be algebraically less than the T. If this assumption is correct in a given hour's comparison an "x" is placed in the column, if it is incorrect an "o" is indicated. Column 14 gives these results comparing the $D1$ figures; Column 15 for the $D2$ figures etc. The last column indicates by a cross a correct judgment as to which of the two crimes was L and which T on the basis of blood pressure. The columns are averaged or totalled as indicated and in the case of the totals the number correct reduced to a percentage of the total number.

The results are inconclusive. The final averages of the D's in the two instances show insignificant differences, pointing in the wrong direction if at all. Not over half of the average D's have the proper sign, and the individual D's are correct in only about half of the cases.

It is impossible to judge which method of computing *D* is the best for none of the three gives more than chance success. There are slight indications that D_3 is the worst method because column 6 shows a clearer tendency than the others toward a truthful *D*, and the following column shows that only 31 per cent. of the L questions are correctly diagnosed on the basis of D_3. This would seem to indicate that after the question is asked and the subject begins to consider his reply, the effect of the lying consciousness dies out rapidly so that during the last breath before replying he is in a practically truthful state of mind.

Neglecting absolute values and attempting merely to select which of the two crimes is L, success is attained in only 2 or 3 of the 7 hours as shown by columns 14 to 16. Nothing depends apparently on the order of the two crimes during the hour for the L came first in 3 of the 7 hours. Of the hours correctly judged the L came first in 2 and the T in the other.

The only suggestion by way of explanation of the failure of series III. is the factor of the habituation, invoked previously with reference to series II. Three of the subjects had already worked in series I. and II. The other subject *Q* was new and he has the largest negative D_1 value on L and the largest positive D_1 value on T in the table. This table, however, is interesting in comparison with that for the next series in which a much different result is obtained when a number of spectators are present. It is to be noted further that the percentage of questions correctly diagnosed is greater in the case of T than in that of L which indicates that the lying consciousness was less frequently present although the person was just as frequently lying.

5. Results of Series IV

The method of Series IV. was identical with that of Series III. except that a number of persons—from 3 to 26—were present watching the subject and trying to determine his guilt or innocence from his reactions. The results are summarized in Table III., which is identical in form with Table II.

TABLE III

Subject	Lie						Truth						Correct Alternative			Correct B.p.	Correct Jury	
	D_1	No. Cor.	D_2	No. Cor.	D_3	No. Cor.	D_1	No. Cor.	D_2	No. Cor.	D_3	No. Cor.	D_1	D_2	D_3		L	T
A	−.29	9/13	−.32	7/13	−.24	5/13	.40	5/7	.40	5/7	.33	5/7	x	x	x	x	8/9	2/9
F	−.16	5/10	−.06	3/10	−.10	5/10	.04	6/11	.15	6/11	.18	6/11	x	x	x	x		
G	−.04	5/11	−.01	4/11	.01	4/11	−.02	5/11	−.10	2/11	−.12	2/11	x	o	o	x	3/5	2/5
K	.05	5/13	−.04	5/13	.05	5/13	.11	9/12	.11	6/12	.07	7/12	x	x	x	x	11/18	7/11
K	.02	4/9	−.02	2/9	.01	2/9	.13	11/12	.01	5/12	.14	9/12	x	x	x	o	18/26	9/26
K	−.07	5/8	−.06	6/8	−.07	5/8	.01	5/13	.24	8/13	.04	7/13	x	x	x	x	1/3	1/3
K	.13	0/7	.17	0/7	.10	0/7	−.04	2/10	.01	3/10	−.12	1/10	o	o	o	x		
M	−.03	4/12	−.01	5/12	.02	5/12	.04	6/8	.22	6/8	.06	5/8	x	x	x	x	7/11	3/11
N	.05	2/9	−.03	3/9	.14	3/9	−.02	1/9	.00	3/9	−.01	3/9	o	x	o	x	7/13	4/13
N	.01	5/12	.03	3/12	.05	3/12	−.14	4/7	−.44	3/7	−.06	4/7	o	o	o	x	4/10	5/10
P	.03	3/8	−.35	4/8	−.28	4/8	.08	6/8	.23	8/8	.10	4/8	x	x	x	x	1/6	3/6
Av. or total	−.03	47/112 42%	−.06	41/112 37%	−.03	41/112 37%	.05	60/108 55%	.08	55/108 51%	.06	53/108 49%	8/11 73%	8/11 73%	7/11 64%	10/11 91%	60/101 59%	36/94 38%

D_1: av. of 3 to 5 breaths between question and answer minus av. of 3 to 5 breaths following answer.

D_2: first breath after question minus first breath after answer.

D_3: last breath before answer minus first breath after answer.

The two additional columns at the right indicate the number of the jurors who judged correctly on a given crime whether the subject was lying or telling the truth. The first figure indicates the number of persons judging correctly and the second figure the total number of persons. In the two cases left blank there were only one or two jurors present.

The results show a considerable relation between breathing and mendacity. The averages of columns 2, 4, and 6 are all negative while those of 8, 10 and 12 are all positive as required by hypothesis. They likewise vary sufficiently from zero to be significant. If the data of columns 2 and 8 are averaged using the original *D*'s for the individual questions the results are − .04 and .05 respectively and the difference between these two figures is 4.4 times the probable error of difference. Considering however, the number of questions in which *D* is of the proper sign there is no greater correspondence than that due to chance. This agrees with the finding in series I that many question show little effect in the breathing but an occasional one does so to a marked degree and thus alters the average value. Disregarding absolute values and attempting merely to select which crime of the two in an hour is L (*i.e.*, has the average *D* algebraically less) the method is successful in 8 of the 11 hours (73 per cent.) using D_1 or D_2, and in 7 hours using D_3.

A brief investigation was made with this series of the factor of variability. The mean variation was found for each of the individual D_1 averages that appears in the second and eighth columns of the table. For example for subject *A* the m.v. of his 13 *D*'s whose average was − .29 was found and likewise for his 7 *D*'s whose average was .40. The question was whether these m.v.'s would be consistently greater on L or vice versa with the various subjects. As a matter of fact in this series 6 of them are greater on L and 5 on T while the average of the former is .14 and of the latter .18. Evidently variability of *D* affords no diagnosis. It did not seem worth while to investigate it in other series.

The columns at the right giving the votes of the jury show the unreliability of judging a person's veracity from his

directly observable behavior. On the average the judgments of L seem to be correct more often than those of T. This reflects a tendency to be suspicious of the subject in this situation and to give an L judgment too frequently. There is little correspondence between the vote of the jury and the magnitude of *D*. Ranking the jury votes with the one with the largest per cent. voting correctly ranked highest and with the D_1 values ranked with the largest negative L highest and the largest positive T highest the correlation for L is .58 and for T — .13 using the method of rank differences squared.

The relative merits of the three methods of computing *D* are shown in Table IV.

TABLE IV

NUMBER OF SIGNS CORRECT.

	D_1	D_2	D_3
L	5/11	9/11	4/11
T	7/11	8/11	7/11
Total	12/22	17/22	11/22
%	55%	77%	50%
$D(T) - D(L)$	.08	.14	.09

The data is taken from Table III. The first column shows for example that 5 of the 11 D_1 averages on L's are correct (*i.e.*, negative in sign), while 7 of the 11 D_1 averages on T's are correct (*i.e.*, positive). The total is 12 out of 22 or 55 per cent. The 0.8 below this represents the difference between the final averages of the D_1 columns in Table III., *i.e.*, .05 — (— .03). This indicates roughly the total extent to which *D* during L differs from that during T. The other columns give similar figures for the D_2 and D_3 methods.

The table shows a tendency for the D_2 method (I/E for breath immediately following question minus I/E for breath immediately following answer) to be the most diagnostic. The value of 77 per cent. as compared with 55 per cent. and 50 per cent., and the corresponding difference of .14 as compared with .08 and .09 point in this direction.

This suggests a return to the data of Series I. to see if a similar relation holds there. It is not worth while presenting

the results in detail, but below is reproduced the "average or total" row from Table I., which was based on *D*1. Beneath these figures are given corresponding ones for *D*2.

	Av. *D* for 16 Questions		No. of Questions *D* Correct		No. of Cards *D* Correct	
	L	*T*	*L*	*T*	*L*	*T*
*D*1	−.05	.01	49%	53%	52%	64%
D2	−.11	−.03	56%	47%	65%	38%

From the standpoint of per cent. of questions or cards having the proper sign for *D* there is little difference. In fact as pointed out earlier this aspect is not diagnostic anyway. From the standpoint of average *D*'s there appears a somewhat greater difference between L and T for *D*2 than for *D*1 although *D*2 for T is slightly negative. This agrees then with the finding in Series IV.

This series, then in which the subjects were questioned regarding an imaginary crime in the presence of a jury, but in which no aggressive cross-examination was given, shows a definite correspondence between the breathing ratio and the telling of truth or falsehood. This correspondence is manifest primarily in the average of a group of questions rather than in the number of individual questions correctly diagnosed. It appears that fully as good if not a better diagnosis can be made using *D*2 (first breath after question and first breath after answer) than using the average of several breaths before and after. In the practical application of such methods it will be desirable to compare a record known to be true with a suspected one. In the present case if the problem is to decide between two records one known to be T and the other L, this is done successfully in 73 per cent. of the cases.

6. Blood Pressure

Although the main interest of the experiment was in the breathing, the systolic blood pressure readings that were taken at intervals tend to substantiate previous blood pressure studies[1] and form an interesting basis for comparison with the breathing.

[1] Cf. Marston, W. M., 'Systolic Blood Pressure Symptoms of Deception,' Jour. of Exper. Psych., 2, 1917, 117–163.

In series I. a judgment was made upon each card on the basis of the b.p. reading after the second and fourth questions. These judgments were made without consulting the record as to the actual nature of the recital. No absolutely fixed criterion could be adopted. The two features that most frequently indicated L were (1) rises of 4 mm. or more above what appeared, on the basis of initial and intermediate readings, to be the normal pressure under the excitement of examination and (2) failure of the pressure to descend to normal from the height to which it had ascended during a previous card.

The last two columns of Table I. show the number of cards correctly judged in this first series. The total shows 71 per cent. correct judgments on both L and T. In series II. (same table) where the problem was to decide whether the first four of the eight cards were T or L, no mistakes were made on the basis of the b.p. This problem was naturally somewhat easier for it was possible to observe the pressure over a longer period and there was a greater chance of catching the subject at the height of a lying consciousness. Evidently the b.p. is considerably more diagnostic than the breathing, and habituation in Series II. is not sufficient to overcome the advantage of more extended observation of the pressure.

In series III. and IV. the problem was to tell which of the two crimes on a given hour was T and which L. In Tables II. and III., the columns headed b.p. indicate by x that this judgment was correct and by o that it was incorrect. In Series III. there are 5 correct judgments out of 7—a higher degree of success than in the case of the breathing. In Series IV., 10 of the 11 judgments are correct, 91 per cent.—again somewhat better than the breathing results.

The question naturally arises as to the correspondence between b.p. and breathing. Out of 64 cards in Series I. the two agree in verdict in 35 cases, *i.e.*, 55 per cent. In Series II. and III. there is very little agreement, but it will be recalled that the breathing results are very unreliable in these series. In Series IV. there is agreement in 7 of the 11 cases, — 64 per

cent. An attempt was made to correlate the two measures in Series IV. as follows. The minimum b.p. reading on a given hour was subtracted from the maximum. The latter always occurred during the L. These differences were ranked. Similarly the breathing *D* for the L on a given hour was algebraically subtracted from *D* for the corresponding T. These values were ranked from the largest positive to the largest negative. These two series of measures were then correlated by the method of rank differences squared. The correlation is $.52 \pm .13$. This indicates a tendency for the person whose b.p. responds readily to the mental state of lying to have his *I*/*E* likewise respond readily in the manner suggested by Benussi.

7. Introspection

The subject was asked at the end of each hour to report upon his mental state during the experiment. He was not asked specific or suggestive questions.

Every subject seemed to have some consciousness of deception. This was sometimes described as an emotion of fear (subjects *A* and *C*) or as a mere 'guilty feeling localized around the heart' (*K*). The sensations of quickened heart beat were noticed likewise by *B*, and there was 'less of this internal disturbance toward the end of the hour.' With *N* there was also fear not of the total situation but of being inconsistent in replies. *N* mentioned tension or strain during lies and relief after the lie was over. *B* mentioned 'excitement like singing in public.' *M* noted 'some sort of resistance during a lie' and he tried to 'build up a sort of mechanism against it.' All of these factors mentioned point to an emotional state or its bodily accompaniments. It is probably futile to theorize until further work has been done upon the emotions as influencing the *I*/*E*, but the following is tentatively suggested. It is possible that fear or excitement on L decreases the *I*/*E* while awaiting the signal to reply and that after the response the ratio returns quickly to something like normal. It is also possible that some such factor as concentration of attention on T while thinking about the

reply increases the I/E with a subsequent return to normal. This factor would likewise be present on L but overbalanced by the emotional factor. If this were true the I/E after both T and L should be somewhat the same. As a matter of fact such is the case. In series IV. the ratios for the breathing following the replies on T and L differ only about half as much as those preceding the replies. If they are compared further with the normal I/E obtained at the beginning of the hour, the ratios after T are almost identical with the normal and those after L are slightly less than normal, as though the emotional effect had not entirely disappeared.[1]

The subjects often described methods by which they tried to deceive the experimenter and jury if the latter was present. In series I. they were instructed to try to fool the experimenter by their manner. This instruction was unnecessary and was subsequently omitted because the situation developed this set anyway. The usual method especially in Series III. and IV. was to try to keep calm or think of extraneous matters during L and to work up a state of emotion or excitement during T. Subject *M* fixed the lie and then forgot about it as far as possible. Sometimes however, he was unable to banish the consciousness that he was lying, as for example in one instance he began thinking what a good alibi he was telling (12 mm. rise in b.p. at this point). He gave especial concern to his mental state at the times of the b.p. readings. He succeeded several times in raising his pressure by general muscular contraction but this was readily detected. Subject *N* used somewhat similar methods. She tried to think of something else—counted or gazed around the room—during b.p. readings on L and she tried to raise her pressure on T by thinking of personal emotional situations or of someone in whom she had 'considerable interest' (presumably of the opposite sex). *Q* (who had discussed the matter with *M*) was quite successful in controlling his b.p. by 'thinking about his girl' during T, and deceived the experimenter, although it was his first appearance in the ex-

[1] Further work is contemplated upon the I/E during various emotional states and during mental work.

periment. It is to be noted however, that the breathing diagnosis was correct. Subject *A* tried the same system and raised his b.p. considerably during T, but told an exciting L which raised it still more.

These efforts (sometimes successful) to influence blood pressure by emotional control are quite significant from the practical standpoint. A guilty defendant who knows that the expert is trying to establish a 'normal' record for comparison with the crucial record may perhaps be able to keep that normal record generally high or to induce abrupt rises in it by means of imagining emotional situations. In this way the normal and crucial records may appear similar and the latter be mistaken for a truthful performance. One partial remedy might be to produce greater emotional reaction during the crucial record by more rigorous questioning or by the natural seriousness of the situation, and thus drive the b.p. to a higher point than that to which it could be raised voluntarily. A better procedure would be to use as many other criteria as possible, notably breathing and some aspect of reaction time. While it is possible to imagine an intense emotion during each reading of the blood pressure it is difficult to do this on every question of an examination. Hence the breathing record should be able to catch the normal points in the performance and combined with the b.p. increase the reliability of diagnosis.

8. Summary

The inspiration expiration ratio (I/E) was recorded with an improved technique while the subject was lying (L) or telling the truth (T). The average I/E for the 3 to 5 breaths following the subject's answer was subtracted from the average I/E for the 3 to 5 breaths between the experimenter's question and the subject's reply. The hypothesis under investigation was that this difference ($D1$) should be negative for L and positive for T. In some instances instead of $D1$, the I/E for the single breath immediately following the subject's answer was subtracted from the I/E for the breath immediately following the experimenter's question ($D2$). Systolic blood pressure was measured at intervals.

The material consisted in the first series of cards containing letters and numbers with L and T trials in a random order. The second series was similar except that the L's at a given sitting all occurred in succession and the T's likewise. The third series involved imaginary crimes with the subject fabricating an alibi or recounting one prepared by an assistant. The fourth series was the same as the third except that a 'jury' was present.

Series I. with simple unemotional material shows some corroboration of the hypothesis. The average of $D1$ for all L questions is $-.05$ while that for all T questions is .04, and the difference between these figures is 4.6 times the probable error of difference. This result is primarily due to a few crucial questions. Many of the D's both L and T are in the vicinity of zero but occasional large negative values are found during L and occasional large positive ones during T. Diagnosis of separate cards in this series (4 questions on a card) is rather unsatisfactory. The results of Series II. are inconclusive, due probably to the habituation of the previous series and the decrease of the deceptive consciousness on similar L cards in immediate succession.

Series III. is likewise inconclusive with regard to the subjects who had performed in the previous series but not with regard to the subject who makes his initial appearance at this point.

Series IV. more nearly approaches practical conditions in material and emotional content. Using $D1$ the grand average difference between T and L is of about the same order as in series I. Using $D2$ it is about half again as large. The effect is here likewise due to a few questions, rather than the majority of questions. In a given 'crime' it is better to note the average D than to note what proportion of the questions have a D pointing in a given direction. The problem of diagnosing the two 'crimes' covered at a sitting, one known to be T and one L, is solved successfully in 73 per cent. of the cases. This is somewhat the type of problem that would be met in practice where a normal record could be obtained for comparison with a crucial one. The use of $D2$

in this series and to a lesser degree in the first seems somewhat superior to the use of D_1. Throughout the work there is generally a little greater accuracy in diagnosis of T than in diagnosis of L. This suggests that the subject is not always conscious of lying.

Systolic blood pressure has a greater diagnostic value than the breathing. In the last series it correctly indicates which of the two crimes is L in 91 per cent. of the cases as compared with 73 per cent. for the breathing. The two criteria correspond appreciably more than half the time and if effort is made to quantify the measures the correlation is around .50.

The lying consciousness appears to have an emotional (probably fear) content and it is possible to influence expressive measurements somewhat by emotional control. Some subjects successfully raise their blood pressure during T by imaginary emotions. It is thus important to have other criteria as a check.

The writer feels that the successful application of psychology to the problems of deception will involve a combination of several methods, including the two above described. Suspects examined by such methods would then fall into three groups,—those in which all measures point toward guilt, those in which all measures point toward innocence and those in which the measures disagree. The middle group would contain with present methods less than half the cases, and this number will decrease with gradual improvement of technique.

AN EXPERIMENTAL STUDY OF MOTOR ABILITY

BY F. A. C. PERRIN

University of Texas

In the investigation now reported, an attempt was made to analyze motor ability. The conclusions advanced are based upon the results furnished by a series of complex motor tests, a series of simple motor tests, an intelligence test, the university grades of the subjects, and character estimates of the subjects. The various measurements recorded were compared and correlated. Since the data for each subject cover a wide range, the study was necessarily intensive rather than extensive. Complete records were obtained, however, from a group of 51 subjects, all of them undergraduate students in the University of Texas. The group was selected on the basis of variety in academic and motor accomplishment. The results reported, therefore, may be assumed to be fairly typical.

I. EXPERIMENTAL—METHODS AND RESULTS

A. Tests and Measurements

Each of the subjects was given individually a series of 17 motor tests. A three-hour laboratory period was occupied in giving this series to each individual, and in recording certain other measurements to be described later. The single period, in spite of its length, was thought to be desirable inasmuch as it guarded against such variations in health and in physical tone as would presumably appear during a longer period of time. And it was found that the novelty of the tests provided against boredom and fatigue in a satisfactory manner. Since the same sequence in giving the tests was followed for all subjects, any possible decline in ability due to the duration of the period was made uniform for all. As a matter of fact, most of the subjects, at the end of the period, were in-

clined to remain in the laboratory room and to experiment further with the tests.

The motor tests consisted of a group of 3, chosen for the purpose of eliciting fairly complex reactions; and of a second group of 14, selected to measure elementary motor functions. A more detailed statement of the distinction between these two groups, and of the meaning of the term, 'motor function,' is reserved for an ensuing section of this paper. The three complex tests were the Bogardus fatigue test, card sorting, and a motor coördination test. The second series of 14 included measurements of reaction time, inhibition of the wink reflex, memory of a motor act, weight discrimination, aiming, the ability to locate points while blindfolded, balancing of various kinds, rhythmic counting, the ability to grasp rhythmic units, tapping, steadiness, tracing, physical strength, and vital capacity. It is obvious, especially in the second list, that the term 'motor' is used in a rather inclusive sense.

1. *Motor Tests*

(*a*) The Complex Motor Tests

1. The Bogardus test was the first of the three given to the subject. Some important alterations in the apparatus were made before it was used. A black wooden cube, somewhat smaller than one of the standard blocks furnished with the apparatus, was glued in the square on which the blocks are ordinarily placed. This served as a platform upon which the subject was forced to place the blocks quickly and accurately as the arms of the machine revolved. The arms, of course, were elevated proportionately. By this device, the task of observing and recording correct reactions was facilitated. While it happened occasionally that a block would fall from the platform, lodge between it and one of the arms, and thus stop the machine, the advantages of the alteration clearly outweighed this disadvantage. In the second place, metal contact points were attached to the revolving vertical tube of the apparatus, and corresponding metal strips fastened to the base, in such a way as to secure automatic registration of the number of revolutions per minute on an electric counter.

These alterations, and the use of an excellent A.C. generator, made possible an almost perfect control over the apparatus.

The subject was given three one-minute tests. His correct reactions were recorded on a hand tally, and their average for the three trials was accepted as the final score. The speed was so adjusted as to demand 54 reactions per minute. A uniform method of explaining the nature of the test, and of demonstrating the technique was followed; but the method of reaching for the cubes and of placing them on the platform was varied, so as to include all legitimate methods of attack. During the actual testing, the task of the experimenter was to observe and record behavior, register successful reactions, and call time. The used of a mimeographed form, the hand tally, and the stop watch made this threefold task not only possible but easy.

The scores of the 51 subjects, in terms of the average number of successful reactions, are given in the following form in order to show conveniently maximum and minimum records, median, modes, and the distribution:

								22														
								22	23	25						30						
		14		16	18		20		22	23	25	26		28	29	30	32					
		14	15	16	18	19	20	21	22	23	25	26		28	29	30	32					
11	14	15	16	18	19	21	21	22	23	25	26	27	28	29	30	32	36	45				

The analysis made of the learning behavior of the subjects proved to be extremely suggestive. It had been made evident during the preliminary work that the subjects varied considerably in their methods of meeting the situations presented by the Bogardus test, and that their variations could be classified and described. A rating chart was accordingly prepared for the present experiment. It provided for the scoring of the subject on his ability to place the cubes accurately on the platform, on his skill in picking up the cubes with his right hand and in depositing them with his left hand, on his general quickness with right hand and left hand respectively, on his smoothness of coördination in using right and left hands simultaneously, on the extent to which he looked in the supply box while reaching for the cubes. It

also provided for the subject's retrospective testimony, solicited at the completion of the three trials, regarding his awareness of these various items of the reaction, with particular reference to his possible conscious or deliberate attempts to change methods. No intimation of the use of this chart was given to the subject.

He was scored on the first set of items by receiving marks of A, B, C, D, and E, representing decreasing amounts of these different abilities. These marks were standardized: a grade of C under any one heading stood for median performance in that ability, or for median employment of that device. The distribution of these marks was not a matter of difficulty. More E's than A's were recorded, but the C's for each item included the middle group.

The results show that of the 27 subjects who attained the median or who exceeded the median score of the number of successful reactions:

21 were above this median in placing the blocks accurately
19 were above in skill with the right hand
18 were above in coördinating the two hands
16 were above in skill with the left hand
16 were above in moving the right hand quickly
15 were above in moving the left hand quickly
10 were above in looking into the supply box

A correspondence between efficiency in performance and skill in the various aspects of that performance of course was to be expected. These results, however, not only indicate what specific aspects of test behavior count for superiority, but, when compared with the data bearing upon conscious change of method, suggest the possibility of the functioning of an intelligence factor. For these aspects are each theoretically capable of receiving conscious attention and emphasis.

But before the results of these additional tests are presented, reference should be made to changes in method on the part of the subjects during the tests. If their reports are to be accepted, these changes were made, in the main, without antecedent deliberation. And furthermore, in nu-

merous instances, they were adopted unconsciously. By this distinction it is meant, in the latter case, that the individual subject frequently found himself in the act of doing something that he had not been doing before, with no memory of the transition, and with no memory of a preceding plan to effect that transition; while in the former case, it is implied that the subject decided on the impulse of the moment to attempt something new. It should be noted that while these shifts in method of attack generally resulted in increased efficiency, they did not necessarily do so. The fact should be emphasized, too, that a few deliberate and conscious variations were made.

The question of the reliability of such introspective evidence of course arises quite readily. But without entering upon an extended controversy, the writer may state that he is very much disposed to accept this testimony at its face value. The introspection called for was not of the dubious sort dealing with fine and possibly theoretical distinctions among mental processes; in fact, it may be straining the term to call it 'introspection' at all. The subject was merely asked to report about his ideas of concrete acts—a task surely within the powers of an intelligent human adult.

The propriety of labeling these methods of attack as intelligence factors is another matter. They are so designated in this article, somewhat arbitrarily, for reasons given above; the subject can attempt, at least, to vary them. And if these attempts are largely unconscious, they are none the less adaptations which result from experience. Skill in placing the blocks is not (necessarily) an affair of intelligence; but the attempt to employ and control skill is a mode of reaction that can be so regarded.

2. In the card-sorting test, ordinary playing cards were used. The subject was required to sort these by suits into four boxes, each plainly labeled by an extra card. He was instructed to hold the shuffled deck, face down, in his left hand; and at the signal, 'begin,' to pick up the cards, one at a time, with his right hand, look at them and put them into the appropriate boxes. He was told not to correct mistakes.

Three sortings were made by each subject, and the average time, registered with a decimal stop watch, for the three was recorded as his final score. Mistakes were charged up against his final score, each one adding to his total time .02 minute. The method of scoring was explained in advance to the subject. The technique, in general was based upon that used by Brown (1).

The use of playing cards, rather than cards bearing original symbols, was decided upon after due consideration had been given to the fact that previous experience might largely determine the score. It was assumed that the subjects were unequally endowed as far as this factor is concerned—not only for card sorting, but for all other tests. An attempt was made to ascertain the extent to which the subjects indulged in card games, and this inquiry did suggest two rough generalizations: that one who does not play is likely to be poor at card-sorting, and that one who does play is not necessarily good at card sorting. But these are fairly weak tendencies. In the absence of any method of measuring transfer, in any of the tests, it was decided merely to recognize this factor as existing in unknown quantities.

A rating chart, similar to the one used in the Bogardus test, but more simple, was employed. It called for ratings upon the following points: helping with the left hand, moving the deck, throwing the cards successfully (as opposed to placing them), holding the deck near the boxes, using the right hand correctly. A vertical yardstick, fastened by a clamp to the experimental table, enabled the experimenter to measure the distance between the deck and the boxes for each subject; but on the other points the letter rating system was used.

The time scores are given in the following table, each column including a range of .05 minute or less:

						1.09						
						1.09	1.05					
						1.08	1.04					
						1.07	1.03					
						1.07	1.03					.85
	1.34		1.24			1.07	1.01		.95	.90	.84	
	1.33	1.28	1.24		1.13	1.07	1.01	.97	.95	.87	.81	
1.47	1.32	1.26	1.23	1.20	1.13	1.06	1.01	.96	.94	.87	.81	
1.44	1.31	1.26	1.22	1.17	1.11	1.06	1.01	.96	.91	.86	.80	

The gradings on the charts establish the fact that the good card sorter must employ certain methods, but they do not prove that these methods alone will guarantee superiority. The efficient individual at card sorting employs his left hand as it is used in dealing cards at a game; throws the cards accurately, and does not attempt to place them in the boxes; picks up the cards with his right hand as he would in dealing; and holds the deck fairly close to the boxes. All of these methods were used by those attaining or surpassing the median score, but they were used also by half of those unable to reach the median rank. Evidently other elements were necessary.

These additional factors were, as one might suspect, speed and the ability to recognize the cards. Instead of attempting to score them on the rating chart, the experimenter tried to measure them by two extra tests. The first, or the speed test, was not used with the present set of subjects, since the test schedule was already over-crowded; but it had been tried out in the preliminary experiment, with a group of 60. It consisted in sorting the cards, irrespective of suit, into the four boxes in order, the remaining conditions of the test being observed. The result was a correlation coefficient of + .80 (P.E., .0605) between this and card-sorting. This coefficient was sufficiently large to establish a tendency for the two activities to correlate.

The ability to recognize the cards (or a corresponding ability) was measured in one of the minor tests—that of reaction time. The Vernier chronoscope was employed, and with it, a set of visual stimuli cards containing single spades, hearts, diamonds, and clubs. Both the simple reactions, and the discrimination and choice reaction tests were given to the subjects. The results of the first yielded a coefficient of + .33 when correlated with card-sorting records; the results of the discrimination and choice test gave + .18. The difference between the two coefficients is probably accidental; together, they would seem to indicate relationship between card sorting and ability to react to card symbols.

The considerations and facts regarding intelligence which

were discussed in connection with the Bogardus test logically call for attention at this point. As in the Bogardus experiment, changes in method were observed by the experimenter, and were reported by the subjects as being both conscious and unconscious adaptations. Due to the fact, however, that speed and recognition seemed to outweigh these methods in importance, no attempt was made to evaluate the intelligence factor by a control experiment.

3. The third and last of the complex tests is referred to as the 'coördination test' in this account. It was essentially a test of the ability of the subject to perform disparate activities simultaneously with his two arms. The apparatus consisted of an aluminum plate upon which were secured a slightly elevated triangle and a corresponding quadrilateral, made of hard rubber. The strips forming these figures were of the uniform breadth of 2 centimeters. The outside circumference of the triangle was 54 centimeters, that of the quadrilateral, 48 centimeters.

The subject was given a stylus for his right hand and a similar one for his left. Either stylus, upon coming in contact with the aluminum plate, would register a sound with an electric buzzer. A metronome was used as a time indicator. At the starting signal, the subject began to trace the figures, both hands going counter clock-wise. He was required to start and to stop both hands simultaneously. The tick of the metronome, which was set at 72, was the cue for the right hand to move from one corner of its figure to the next; and for the left hand to do the same. The right hand therefore described its figure (the quadrilateral) in four counts, while the left hand completed its round in three.

In this test, a practice period of one minute was the unit of time. It was followed by a minute of rest. The criterion of successful accomplishment was the ability to trace both figures without error for 10 successive rounds, with the quadrilateral as the counting standard. The number of one minute practice periods required for this was the final score.

On the whole, the task was a difficult one. The majority of the learners agreed that it was the most difficult of all

the tests, and two of them (both men) passed it up as impossible. For some, however, the coördination came quickly and easily.

The scores show an extremely wide distribution, but with a number of superior scores at the extreme excellent end. In the table, fractional parts of a minute are omitted; and the two subjects who failed are represented by question marks. Since one of them persevered longer than the other, he was assigned, possibly with poor logic, to the 'next poorest' position. The scores are as follows:

| | | |

and-choice reaction. Four series of stimuli, therefore, were given.

The inter-correlations existing among the four series, and the correlations between these and the results of other tests, will be presented later. The following condensed table discloses the tendencies found in the results:

	Longest Reaction Time	Shortest Reaction Time	Median Reaction Time
Simple reaction: colors	460 σ	190 σ	298 σ
Simple reaction: cards	474 σ	180 σ	312 σ
Discrimination-and-choice: colors	598 σ	236 σ	336 σ
Discrimination-and-choice: colors	540 σ	244 σ	390 σ

2. The wide range of meanings of the term 'inhibition' made the selection of a test of that function, or group of functions, a matter of considerable difficulty. At the same time, inhibitions of some kind presumably appear in any motor learning process. It was thought advisable, therefore, to introduce some test of this nature in the series. The attempt to control the wink reflex seemed to afford an unambiguous example of inhibition in one accepted sense of the term, and that ability was accordingly selected, with some misgivings, as a representative of the function.

The apparatus employed was constructed in the laboratory. A section of a piano movement was altered to suit the purposes of the test, by the addition of a square of thick glass, and by the construction of two cork hammers, which could be made to strike the glass by the operation of two levers. The person to be tested rested his forehead on the glass and fixated upon a designated mark. His position was such that the right hammer would strike immediately in front of his right eye, and the left hammer in front of his left eye.

The subject was given 48 trials. It was explained to him that in any one of these he might be tested by the right hammer, by the left, or by both together. He was scored on one of three points in each trial—on complete success, on partial success, or on absolute failure. These scores were

weighted respectively as 2, 1, and 0; and the sum of his weighted scores constituted his final score. As the table of scores indicates, the most successful of the subjects made a score of 85 (out of a possible 96) while eight of them were unable to inhibit at all.

0
0
0
0
0 2 4
0 2 4 30
0 1 2 3 4 18 19 21 30 36 41
0 1 2 3 4 7 10 12 13 17 18 19 20 21 22 24 30 36 38 39 41 44

48
48 54 58 63 72 85

3. The point has been made that the three major tests as conducted were actually tests of the acquisition of skill. This means necessarily, that in each of the three the factor of habit formation was involved. Since this in turn implies the fixation of movements through repetition, it was thought desirable to measure this function.

A pencil maze with no *culs-de-sac* was constructed for this purpose. It contained 28 segments, with a total length of 313 centimeters. The subject was blindfolded, and was told to trace the path-ways in the maze three times, from entrance to exit, using a blunt rubber tracer. He was instructed, in addition, to memorize the sequence of paths; but he was warned not to do any retracing. The accuracy of his drawing, made from memory at the completion of the third trial, was the measure of his success.

At the best, it can be claimed that this apparatus, as employed, measures the immediate memory for a specific sequence of motor acts. It does not necessarily measure kinaesthetic memory: the subject was at liberty to employ articulatory processes, visual imagery, or any other actual or fictitious mental device. The same can be said, of course, of his methods of establishing habits in the complex tests.

The graphs of the subjects were ranked by four judges,

on the basis of the number of segments correctly reproduced, minus the number of segments interpolated in any false fashion. With the exception of the five poorest drawings, which were exceedingly wild and inaccurate, the graphs were easily arrayed in a series of approximately equal gradations of merit. The best graph was an almost accurate drawing of the maze path.

4. The weight descrimination test was given on the rather thin supposition that it measures some phase of motor discrimination likely to enter into more complex processes. In general, the method described in Whipple's 'Manual' was followed. The subject made 12 trials, and gave a judgment for each. The possible range, therefore, was from 0 to 12; but the actual range was from 6 to 12, as the table of scores indicates:

			9		
			9	11	
			9	11	
			9	11	
			9	11	
			9	11	
			9	11	
	7		9	11	12
	7	8	9	11	12
	7	8	9	11	12
	7	8	9	11	12
6	7	8	9	11	12

5. In the aiming test, as in the tracing and steadiness tests, the Stoelting apparatus was used and the general method described by Whipple was followed. Records were obtained for both the right and left hands. The sum, rather than the averages, of the distances between points of intersection on the target blank and pencil points was taken as the unit for scoring, since this permitted the experimenter to rank subjects in order of merit. Since there were 10 points of intersection on the target blank, the figures in the tables below are centimeters and decimal parts of centimeters. Any one score, again divided by 10, would give the average number of millimeters for the individual between pencil point and intersection point.

									4.1	
					8.1	7.0			4.2	
					8.2	7.2	6.0	5.3	4.2	
					8.3	7.5	6.1	5.5	4.5	
				9.0	8.3	7.5	6.2	5.6	4.6	
			11.5	10.4	9.0	8.3	7.6	6.2	5.7	4.8
		12.0	11.6	10.7	9.5	8.4	7.6	6.3	5.8	4.9 3.4
		12.3	11.9	10.7	9.7	8.9	7.9	6.3	5.9	4.9 3.5

Distribution of scores—right hand.

20.3	18.8	17.4	16.3	15.5	14.2	

					10.0					
					10.0					
					10.2					
					10.2					
					10.3	9.1		7.4	6.0	
				11.3	10.3	9.1		7.5	6.4	
				11.4	10.4	9.3		7.6	6.5	5.1
				11.5	10.4	9.3	8.2	7.6	6.5	5.3
				11.5	10.8	9.5	8.5	7.7	6.7	5.5
13.6	12.2	12.0	11.7	10.9	9.5	8.8	7.7	6.8	5.8	4.6

Distribution of scores—left hand.

6. The aiming test was followed by one in which the subject was directed to register a point with a pencil with his eyes open, and to make a second one, with his eyes closed, as near the first as possible, after having in the interim dropped his hand to his side. The base board used in the aiming test held the paper in place, and the subject stood at the same distance from it observed in the previous test. Three trials were made for the right hand, and three for the left; and the sum of the distances, as before, was recorded as the subject's score.

This test is reminiscent of the ancient parlor game of pinning the tail on the donkey. It was introduced chiefly for the purpose of comparison with the aiming test, since one of the central problems of the experiment was to describe, if possible, the degree of modality of the motor functions.

In the appended table, the scores represent sums of three trials, in terms of centimeters and decimal fractions thereof:

				8.1				3.0			
				8.2		6.1		3.1			
				8.3		6.1		4.0	3.3		
			10.0	8.6		6.2		4.1	3.3		
			10.4	9.0	8.7		6.3	5.2	4.2	3.5	
12.1	11.1	10.7	9.3	8.8	7.1	6.5	5.5	4.2	3.5		
12.2	11.7	10.9	9.3	8.8	7.6	6.6	5.6	4.2	3.8		
12.3	11.8	10.9	9.3	8.9	7.9	6.8	5.9	4.9	3.8		

Distribution of scores—right hand.

					8.0			5.1			
					8.0			5.1			
				10.1	9.0	8.1		6.0	5.2		
				10.3	9.1	8.2		6.0	5.3		
				10.3	9.3	8.3		6.1	5.5		
				10.7	9.8	8.7		6.2	5.5		3.0
		13.0	11.3	10.7	9.8	8.9	7.6	6.2	5.6	4.7	3.2
17.5	14.4	13.9	12.4	11.7	10.8	9.9	8.9	7.7	6.4	5.7	4.7	3.4

Distribution of scores—left hand.

7. In the balancing tests, five different things were demanded of the subject. He was asked to walk forwards, and backwards, on the edge, $\frac{1}{2}$ inch in width, of a 12-foot board, held up by blocks nailed to its sides; and he was required, in addition, to take a few steps on this board, turn, and walk backwards. He made five attempts to perform each of these tasks. He was then required to balance himself, first on his right foot, and then on his left, on a platform, 10 by 4 inches in size, fastened on the end of a 2″ x 4″ plank, 8 inches long, which was placed on the floor in a vertical position. And finally, he was directed to balance a strip of wood 29 inches in length, 1 inch in width and in thickness, vertically, first with his right forefinger, and then with his left.

For the 'walking and turning' tests, errors were counted; and for the 'balance—standing' test and the 'balancing stick' test, time was counted. The errors were weighted, since the three things to be done were not of equal difficulty. The result of this was a range of from 0 to 90 possible errors. A maximum time limit of three minutes was called for in the balancing tests.

The following results give some indication of the difficulty experienced by the subjects in the various tests:

	Lowest Score, Greatest Number of Errors	Highest Score, Least Number of Errors	Average Number of Errors
Walking and turning tests.....	56	9	32
	Lowest Score, Shortest Time	Highest Score, Maximum Time (3 Subjects)	Average Time
Balancing—right foot.....	.12 minute	3 minutes (9 subjects)	.45 minute
Balancing—left foot.....	.18 minute	3 minutes (4 subjects)	.53 minute
Balancing stick—right hand.....	.03 minute	3 minutes (2 subjects)	.09 minute
Balancing stick—left hand.....	.02 minute	3 minutes	.07 minute

8. Inasmuch as the reactions in the Bogardus and coordination tests were of necessity rhythmic, it was surmised that tests of this function might possibly be of value. In the first of two tests devised for this purpose, the subject counted aloud with a silent pendulum, while observing it, for five counts. He then, having had opportunity to catch the rhythm, turned, closed his eyes, and attempted to continue counting aloud at the same rate. He was stopped when he lagged behind or forged ahead one count. The number of counts constituted his score.

The scores were fairly well distributed, and were rather widely scattered, as the table shows:

```
      13      17
      13 15   17          23 24    26 27    29 30
10 11 13 15 16 17 19 20 22 23 24 25 26 27 28 29 30

32
32          37                          60
32 33 34 36 37 40 41 43 46 47 51 53 57 60 66 75 84 93
```

9. The second of the rhythm tests was designed to measure the subject's ability to grasp rhythmic units. Eight series of dots and dashes were printed upon a sheet of paper, and this paper was fastened on the drum of a kymograph. This instrument was placed in a horizontal position. As the drum revolved, at a slow rate, and carried the dots and dashes constituting one series past a stationary mark, they were translated into sounds made by an electric buzzer. The subject, after hearing one complete series, gave his verdict of

'yes' or 'no,' according to his judgment of whether or not the series of sounds came in rhythmic groups. Four of the series were arranged to show these groups, and the number of units in these four varied from 5 to 9. The range of possible correct judgments therefore was from 0 to 8. While this possibly permitted too much of the chance factor, the subjects on the whole were quite certain of their judgments. The scores were distributed as follows:

```
                5
                5
                5 6
                5 6 7
                5 6 7
                5 6 7
                5 6 7
                5 6 7
                5 6 7
                5 6 7
              4 5 6 7 8
              4 5 6 7 8
          2 3 4 5 6 7 8
          2 3 4 5 6 7 8
```

10. The Burrough's Calculator was substituted for the tapping board in the speed test. This instrument is at least more foolproof than the board. The subject held a pencil, rubber end down, and used the lower, right-hand '1.' He was given three trials of .10 minute each, and his average for the three was his score. Right-hand and left-hand records were obtained, as follows:

```
                      40
                      40          43 44
                      40          43 44
                      40          43 44          47
                38 39 40 41       43 44          47 49
          35 36 38 39 40 41 42 43 44 45 46 47 49    51
29 32 35 36 38 39 40 41 42 43 44 45 46 47 49 50 51 53
```

Distribution of scores—right hand.

```
                                  42
                      39          42
                      39          42
                      39          42
                      39 40 41 42
             35       38 39 40 41 42    44
          34 35 36    38 39 40 41 42 43 44 45 46 48 49
30 31 33 34 35 36 37 38 39 40 41 42 43 44 45 46 48 49 51
```

Distribution of scores—left hand.

11. The subjects were tested for steadiness, both for the right hand and for the left. The Stoelting apparatus was used, and the Whipple directions followed. Since time was recorded with the decimal stop watch, the scores read in terms of minutes and decimal parts of a minute:

															.27			
					.13	.14	.16				.23		.26	.27		.31		
.06	.08	.09	.11	.13	.14	.16	.17	.19	.20	.21	.23	.24	.26	.27	.30	.31		
					·45													
·32	·33				·45								·74					
·32	·33	.34	.39	.44	.45	.46	.48	.50	.55	.60	.69	.70	.74	.80	.98	1.01		
1.32	1.77	2.73																

Distribution of scores—right hand.

		.08	.09		.12	.13		.15		.18				.26		.28		
.05	.07	.08	.09	.11	.12	.13	.14	.15	.17	.18	.19	.23	.25	.26	.27	.28		
			·33															
.30	.31	.32	.33	.35	.37	.41	.42	.43	.44	.50	.55	.58	.64	.66	.77	1.03		
1.11	1.55	2.90																

Distribution of scores—left hand.

12. In the tracing test, again the Stoelting apparatus was used, and the Whipple instructions followed. The board was kept in but one position, however, and the movement of the subject's hand was in one direction only—towards his body. Records were taken for the right and for the left hands. They are given in the table in terms of centimeters:

		33					38											
23·5	28	33	34	34·5	37·5	38	41	42·5	43·5	44	44·5	45·5	47	47·5	48	49		
	50		51				55		57		58·5		60·5					
49·5	50	50·5	51	53	53·5	54	55	56	57	57·5	58·5	60	60·5	61	61·5	62		
	65																	
64·5	65	65·5	67·5	68·5	69·5	70	70·5											

Distribution of tracing scores—right hand.

														50·5				
35					40					48			49·5	50·5				
35	35·5	38·5	39·5	40	42·5	44·5	45	46	48	48·5	49	49·5	50	50·5	51	52·5		
	53·5		55															
53	53·5	54·5	55		56·5							63			64·5			
53	53·5	54·5	55	56	56·5	57	59·5	60·5	61	61·5	62	63	63·5	64	64·5	65		
66	68·5	71																

Distribution of tracing scores—left hand.

13. Physical strength was tested for the right hand by means of the Smedley dynamometer, and for the back and legs by the back and leg dynamometer. In order to get a uniform distribution for the group of boys and girls, weights were assigned to the girls' records: the difference between the median record of the boys and the median record of the girls was added to the score of each girl. The weighted scores are given in the tables:

42 47 49 50 54
42 47 48 49 50 52 54 56
42 44 46 47 48 49 50 51 52 54 56 58 61
24 31 32 34 35 40 49 42 44 46 47 48 49 50 51 52 54 55 56 57 58 59 61 65

Dynamometer scores—right hand.

87 95 97
77 87 95 97
77 80 86 87 90 93 95 97 99
60 67 77 80 82 85 86 87 88 89 90 92 93 95 97 99 100 101 107 108

112
110 112 119 120 124 125 137 158 159 162 205 212

Dynamometer scores—back strength.

142
100 120 130 142
50 94 96 100 106 110 120 122 128 130 138 140 142 143 148 155 156 158

160
160
160
160
160 170 174 202
160 162 164 170 174 175 176 177 180 188 196 198 202 212 220 225 230 253 260

Dynamometer scores—leg strength.

14. The 'vital capacity' test was the last given to the subject. The wet spirometer was employed, and its cubic centimeter readings recorded. Again the records of the girls were weighted, after the manner described in the preceding test. By this method, 15 was added to the record of each girl. The final results are given in the table:

| | | |

											2.78		
											2.73		
											2.72		
								2.39			2.72	2.88	
								2.38		2.55	2.67	2.72	2.86
							2.27	2.37		2.54	2.63	2.72	2.85
1.14	1.67	1.78	1.82	2.00	2.15	2.23	2.32	2.47	2.52	2.62	2.70	2.81	
							3.79						
2.98	3.17				3.55		3.79						
2.97	3.13	3.28			3.54		3.77						
2.96	3.11	3.23	3.33	3.42	3.52	3.64	3.74	3.89	4.14	4.17	4.43		

4. *Estimates of Character*

Although the plan to secure estimates of character for each of the subjects was part of the original program, the specific items enumerated on the chart were formulated entirely as the result of the observations of the behavior of the subjects during the motor testing period. The chart was mimeographed, and copies of it were given to associates of the subjects. Care was taken to see that the judges did not compare notes before recording their ratings. To the extent that it was possible, the same judges were used for the different subjects; but since not more than half of them, at best, knew intimately more than a dozen of the subjects, this could not be effected with any degree of consistency. All ratings were made under the immediate supervision of the writer. The judges were urged to be quite honest, on the promise that their names would not go on record, and that their unsigned charts, even, would not be seen by anyone except the experimenter.

Five ratings on all items listed were secured for each subject, and the median rating was taken as the final one. For purposes of making the idea clear to the subjects, the traits were arranged by opposites. The specific opposed traits used can easily be criticized; indeed, they were criticized by many of the judges. But they did seem to represent the contrasted traits observed during the laboratory testing. The judges were asked to place the subject in the highest 10 per cent. of University undergraduate students, the next highest 10 per cent. etc., down to the lowest 10 per cent. The complete chart is here reproduced:

Please grade ———————————————— on the following traits, by placing him in the highest 10%, the second 10%, the third 10%, etc., of University students. Try to avoid either over-estimation or under-estimation. These scores will be used only for purposes of correlation, *i.e.*, individual records will not be used. Place a check (V) to indicate in what percentage group the individual stands in your estimation.

Your judgments will be regarded as absolutely confidential.

	Highest 10%	Intermediate								Lowest 10%	
		9	8	7	6	5	4	3	2		
Poise, self-possession											Embarrassment, resulting in incapacity
Vivacity, energy, 'pep'											Entire absence of such traits
Determination, ability to 'see it through'											Indifferent, willing to let things slide
Practical, common sense, general intelligence											Stupidity, absence of common sense
Aptitude for social or group leadership											Easily influenced by any one
Likes to do things well for the sake of the things themselves											Motives extraneous —money, praise, grades
Sense of rivalry, desire to excel											Perfectly willing to be the most inferior
Practical											Visionary
Deliberate											Spontaneous
Conceited											Depreciative
Optimistic											Pessimistic
Frank, sincere											Posing, insincere
Serious											Frivolous
Graceful, skilful											Awkward, clumsy
Consistent in traits											Variable in traits
Skill in tennis, swimming, etc.											

The ratings for the nine traits which seemed to be of the greatest significance were plotted against the rankings in the Bogardus, coördination, and card-sortings tests respectively. The scatter diagram was used for this purpose, and the actual median of the group represented on the diagram was used. This median, in the case of the character ratings, was rather high, indicating a tendency for these ratings to favor our group of 51 subjects on desirable traits; but as the judges frequently pointed out, this group was a selected one. Ap-

proximately 50 per cent. of the subjects were placed in the highest 30 per cent. of desirable traits. This is probably a fair distribution, when we consider the fact that the subjects were of sophomore standing or better, although they were compared, in the ratings, with the entire undergraduate body.

B. Correlations and Comparisons

The principal measurements obtained from the various tests and rating charts were compared and analyzed by three methods: (1) coefficients of correlation among the test scores were ascertained; (2) scatter diagrams for the character estimates and tests were plotted, and their results expressed in tabular form; (3) a table was constructed for the purpose of showing the variability of the subjects in five of the more important measurements. The results of the first two methods are submitted as showing general tendencies, or lack of tendencies, only; the results of the third method may help to explain some of these tendencies.

1. The coefficients of correlation obtained, and their (approximate) probable errors, are given in the following table:

	Bogardus		Card Sorting		Coördination	
	Coefficient	P. E.	Coefficient	P. E.	Coefficient	P. E.
Complex Motor Tests						
1. Bogardus					−.10	.0944
2. Card Sorting	−.06	.0949			+.36	.0834
Simple Motor Tests						
1. Reaction time	+.16	.0932	+.21	.0913	−.06	.0949
2. Inhibition	+.02	.0952	−.11	.0943	+.02	.0952
3. Motor memory	+.02	.0952	+.22	.0908	+.17	.0926
4. Weight discrimination	−.07	.0947	−.03	.0951	−.01	.0953
5. Aiming	+.24	.0898	−.09	.0945	+.12	.0943
6. Aiming—blind-folded	+.29	.0873	+.05	.0949	+.04	.0950
7. Balancing	+.38	.0813	−.06	.0948	−.11	.0943
8. Rhythm—pendulum	+.17	.0926	+.20	.0915	+.29	.0873
9. Rhythm—kymograph	−.02	.0952	+.12	.0943	+.12	.0942
10. Tapping	+.22	.0908	+.02	.0952	+.01	.0953
11. Steadiness	+.09	.0945	−.04	.0950	+.01	.0953
12. Tracing	−.11	.0943	−.01	.0953	+.09	.0945
13. Dynamometer	−.22	.0908	+.20	.0915	+.39	.0807
14. Vital capacity	−.14	.0939	+.09	.0945	+.22	.0908
15. Army Alpha test	+.03	.0951	+.02	.0952	+.10	.0944
16. University grades	−.17	.0926	+.01	.0953	+.21	.0913

2. The results of the scatter diagrams were formulated

in terms of the number of subjects falling below the median in one series of measurements who also fell below in a second series; and of the number who ranked above the median in the two series. This seemed to be the best method of comparing the results actually obtained. With a greater number of cases, and with better correlations, finer comparisons could be made. An attempt was made to discover the critical scores, after Thurstone's method, but this was possible only in a few of the diagrams (3). The results are as follows:

(*a*) *Bogardus and Character Estimates*

Of those below the median in Bogardus,		Of those above the median in Bogardus,	
16 (31%)	were below the median;	14 (27%)	were above the median in Poise
17 (33%)	" " " "	15 (29%)	" " " " " Vivacity
18 (35%)	" " " "	17 (33%)	" " " " " Determination
12 (23%)	" " " "	13 (25%)	" " " " " Common Sense
12 (23%)	" " " "	10 (19%)	" " " " " Interest
11 (21%)	" " " "	13 (25%)	" " " " " Rivalry
14 (27%)	" " " "	11 (21%)	" " " " " Deliberation
16 (31%)	" " " "	14 (27%)	" " " " " Grace
9 (17%)	" " " "	11 (21%)	" " " " " Skill

(*b*) *Card Sorting and Character Estimates*

Of those below the median in Card Sorting,		Of those above the median in Card Sorting,	
17 (33%)	were below median;	17 (33%)	were above median in Poise
14 (17%)	" " "	12 (23%)	" " " " Vivacity
20 (39%)	" " "	19 (37%)	" " " " Determination
8 (15%)	" " "	8 (15%)	" " " " Common Sense
17 (33%)	" " "	16 (31%)	" " " " Interest
14 (27%)	" " "	15 (29%)	" " " " Rivalry
12 (23%)	" " "	13 (25%)	" " " " Deliberation
14 (27%)	" " "	13 (25%)	" " " " Grace
10 (19%)	" " "	9 (17%)	" " " " Skill

(*c*) *Coördination and Character Estimates*

Of those below the median in Coördination,		Of those above the median in Coördination,	
13 (25%)	were below median:	14 (27%)	were above median in Poise
16 (31%)	" " "	14 (27%)	" " " " Vivacity
10 (19%)	" " "	8 (15%)	" " " " Determination
9 (17%)	" " "	10 (19%)	" " " " Common Sense
14 (27%)	" " "	13 (25%)	" " " " Interest
15 (29%)	" " "	16 (31%)	" " " " Rivalry
12 (23%)	" " "	13 (25%)	" " " " Deliberation
12 (23%)	" " "	11 (21%)	" " " " Grace
16 (31%)	" " "	13 (25%)	" " " " Skill

(d) Complex Motor Tests

Of those below the median in Bogardus,	9 (17%)	were below median in	Cards
" " above " " " "	11 (21%)	" above " " "	
Of those below the median in Bogardus,	9 (17%)	were below median in	Coördination
" " above " " " "	10 (19%)	" above " " "	
Of those below the median in Cards,	16 (31%)	were below median in	Coördination
" " above " " " "	15 (29%)	" above " " "	

(e) Complex Motor Tests and Army Alpha Test

Of those below the median in Bogardus,	11 (21%)	were below median in	Alpha
" " above " " " "	12 (23%)	" above " " "	
Of those below the median in Cards,	15 (29%)	were below median in	Alpha
" " above " " " "	14 (27%)	" above " " "	
Of those below the median in Coördination,	15 (29%)	were below median in	Alpha
" " above " " " "	13 (25%)	" above " " "	

(f) Complex Motor Tests and University Grades

Of those below the median in Bogardus,	10 (19%)	were below median in	grades
" " above " " " "	10 (19%)	" above " " "	
Of those below the median in Cards,	15 (29%)	were below median in	grades
" " above " " " "	14 (27%)	" above " " "	
Of those below the median in Coördination,	17 (33%)	were below median in	grades
" " above " " " "	16 (31%)	" " " " "	

3. The coefficients of correlation obtained establish either poor correspondence or absence of correspondence between any two series of measurements. The scatter diagrams indicate that a slight positive relationship exists, inasmuch as an individual who falls below median in any one measurement stands a certain chance of falling below in any other series. Neither of these devices, however, gives us information regarding the number of subjects responsible for the low correlations. It is highly important to know whether they were caused by a few subjects showing extreme fluctuations from test to test, or by all of the subjects showing a consistent and uniform variation.

A number of tabulations were made for the purpose of ascertaining the nature of the variations made. The results of two of these may be presented briefly:

(*a*) In one table, the rank of each subject in each of the 19 tests, exclusive of the rating charts, was determined and stated. According to this table, three subjects ranked above the others. The median rank, for each of the three, was 18:

their rankings were above this in 9 of the tests, and below it in the remaining 9. The poorest subject had a median rank of 32. In half of the 19 tests, he ranked below the thirty-second position. The distribution of these ranks was fairly symmetrical. Evidently, when all of the measurements are taken into consideration, the fluctuations in the record of any one subject are extreme.

(*b*) In this table, however, it is assumed that all tests, mental and motor, are of equal value. But it is obvious that some of the complex tests, although not necessarily more difficult than the simpler ones, clearly are of greater significance so far as the kind of ability which they call for is concerned. Accordingly, a selection was made of the five tests which seemed to measure the most highly organized forms of behavior ability, and the records of all of the subjects were compared, with reference to their positions above and below the median lines. The five selected were the Bogardus, card sorting, and coördination motor tests, the Army Alpha test, and university grades. The results are as follows:

(i) None of the 51 subjects were above median in all of the five.

(ii) 11 of the subjects were above median in 4 of the five measurements. Each of the 11 subjects, therefore, fell below in one of the 5. The low scores were distributed as follows: Bogardus, 5; coördination, 3; cards, 1; Army Alpha, 1; university grades, 1.

(iii) 14 subjects were above median in 3 of the 5 measurements. Each of the 14 fell below in 2, giving 28 low scores, as follows: Bogardus, 7; coördination, 7; cards, 5; university grades, 5; Army Alpha 4.

(iv) 10 subjects were above in two of the tests. The 30 low scores were due to university grades, 7; coördination, 7; Alpha, 7; cards, 6; Bogardus, 3.

(v) 8 subjects were above in one test only. Among these 8, the following scores obtain: University grades, 8; Army Alpha, 8; coördination, 6; Bogardus, 5; cards, 5.

(vi) 5 of the subjects were below the median in each of the 5 tests.

II. DISCUSSION

The primary purpose of this investigation was an analysis of motor ability. It is sufficiently obvious that any conclusions based upon the results obtained must be regarded as tentative and even speculative. A reason for this statement, even more pertinent than the small size of n, is to be found in the fact that an explanation of motor ability involves a satisfactory account of the entire field of human reactions. A number of interpretations or theories, however, were suggested by the results; and our purpose must be satisfied with an examination of these various possible interpretations, with such an evaluation of each as the evidence seems to warrant.

A brief enumeration of these theories will serve to clarify our discussion. Motor ability may be regarded as (*a*) a general factor or function (*b*) the resultant of a number of specific functions, operating simultaneously (*c*) an ability based upon a few general modes of motor reaction (*d*) a function of intelligence (*e*) a response which is determined by more general traits of character, rather than by motor habits or abilities (*f*) a complex reaction which does not permit analysis in terms of functions or general methods of response. While this list of possibilities is fairly comprehensive, it does not extend beyond the interpretations suggested by the data.

A. It seems quite patent that motor ability is not general, but that it is somewhat definitely specialized. As a statement of fact, this means merely that absence of correlation obtained among the scores of the various tests, both complex and simple. Similar results were obtained long ago by Wissler (4). But this investigator was not especially interested in the analysis of motor ability, and his tests consequently were not primarily motor. To the extent that the two investigations are comparable, the results of Wissler are corroborated by the present findings.

To this statement of fact must be added one of explanation. A factor of general motor ability may exist; but if it does, it is lost in combination with other (non-motor) factors

as the individual proceeds from one motor test to another. The futility of attempting to predict the degree of success in the coördination test from the record established in the Bogardus test, fully demonstrated in the absence of correlation between the two, was made quite plain to the experimenter before the results were tabulated.

The analogy with the concept of general intelligence is of interest in connection with the foregoing conclusion. Of course, we could pool results, as psychologists have done with mental test results, and assume that the final score of the subject would represent his general motor ability. And we would compare the A.D.'s in motor test scores with the A.D.'s in mental test results. But if this procedure resulted in a marked difference in the two sets of A.D.'s, it would show merely that a factor, or general set of factors, is more operative in the one than in the other; it would not demonstrate the presence of a single general factor, either of intelligence or of motor ability. Undoubtedly, motor tests could be selected so homogeneous in character that high correlations would obtain among them. Varieties of card-sorting tests would be a case in point. Our motor tests however, were selected on a priori grounds, with the end in view of testing each of the possible (assumed) motor functions, while mental tests are frequently assembled on the basis of high correlations, either among themselves, or with school grades. In our conclusion, then, it is not implied that motor ability is to be distinguished from mental ability by the fact of its greater variability.

B. Our data, on the whole, argue against the theory that a complex motor performance can be explained in terms of a number of unit motor functions. It would assuredly simplify psychological doctrine if we could so explain complex motor ability as an agregate, or as a resultant, of elementary motor habits, reflexes, or the like. This would give us an atomic concept which would permit easy analysis of complicated responses, expressed graphically by means of a motor function profile, or some similar device. But motor skill is seemingly not so simple.

This conclusion is based, first of all, upon the fact that no complex test correlates sufficiently with any simple test, or group of simple tests, to justify such an assumption. An attempt was made to apply the method of partial correlations, with the scores which promised reward for the work involved, but no positive results were obtained. The differences ascertained among the individual coefficients of correlation are small—so small that, in the majority of cases, we may believe them to be chance differences. It is true that differences extending over 25 points (about twice the P.E. of the difference between two coefficients) may be assumed to be of significance; but when we note the coefficients which are relatively high, we find that the tests which yielded them fail to show theoretical relationship. Thus, among the relatively high coefficients are the following: motor memory and card sorting; balancing the body and Bogardus; physical strength and coördination; speed and Bogardus. Of these, only the last named seems to suggest logical relationships. And, although we might reasonably expect an actual relationship between card-sorting and speed, coördination and tracing, none was made evident by the coefficients. Even card-sorting, and the discrimination-and-choice reaction test in which card symbols were used, gave a coefficient of only + .18. In fact, the distribution curve of the coefficients was proved to be approximately normal, suggesting that the differences were due to chance factors.

Even more to the point is the question of the existence of simple motor functions. Notwithstanding the fact of objective similarity in the various manifestations of speed, of rhythm, of balancing, and of reaction, it is doubtful if these rubrics represent true functional entities. We can believe, of course, without question, that physical strength is, generally speaking, physical strength; and that a capable piano mover might qualify at the task of loading railroad ties. But it does not follow that the speed required in base ball is the speed demanded in typewriting, or that the rhythm necessary in dancing is that involved in the coördination test.

In connection with this conclusion, the following coefficients are of interest:

Simple color reactions and discrimination-and-choice color reactions, + .59
Simple color reactions and discrimination-and-choice card reactions, + .40
Simple color reactions and simple card reactions, + .33
Locating a point (Test 6) and aiming, + .26
Speed in tapping, and speed in simple color reactions, + .06
Rhythm (units) and Rhythm (counting), + .04

The principle of identical elements is suggested by the above list of coefficients. But this principle should not be confused with assumption that speed, for instance, is a function which can be measured and tested. The criticism may be made that the simple motor functions here described were not the ones which actually constituted the three major test reactions, or that the tests employed were not adequate tests of the functions operative. The ones described, however, seemed at the time to enter into the complex performances. They are, moreover, the ones ordinarily described in our literature. And it is possible that some of the 'simple' functions were, in reality, as complex as any of the three major tests. We can say merely that they were 'simple' in terms of the number and scope and variety of observable movements involved. Only to the extent that these movements can be transferred to a number of other tasks can they be regarded as functions.

C. A third explanation remains, still within the field of motor reactions. It is conceivable, at least, that an analysis of motor skill will disclose, not a general function, and not a sum of specific functions, but a series of general modes of reaction to be characterized as longitudinal rather than transverse. Thus, we may assume a common ability in gross reactions as opposed to fine; or uniform skill in a series of tasks depending upon the amount of ambidexterity involved; or an aptitude in tests involving some form of locomotion as over against some form of hand manipulation. The list, of course, could be extended considerably.

Unfortunately, our experimental data do not contribute directly to an evaluation of any theory of this type. Some attempt was made to isolate the factor of ambidexterity, but the evidence obtained was not conclusive. It does show, however, that the degree of ambidexterity possessed

by an individual seems to vary from test to test. This particular explanation principle, therefore, promises little. Again, a test involving fine coördinations (the game of jack straws) was given to a group already tested with the Bogardus apparatus, but the results failed to indicate group differences. The explanation suggested, then, must be regarded as hypothetical only.

D. According to a fourth theory, motor ability may be regarded as a function of intelligence. All of the tests demanded perceptual analysis of a fairly complex order. Most of them provided for some variation in the method of attack, thereby permitting the subject to exercise a degree of choice in his reactions. Moreover, since successive scores were recorded, each test assumed the proportions of a learning process; and each involved, in consequence, a play of selective and retentive factors. Verbal processes, analytical as well as expletive, were elicited in abundance from the subjects. Theoretical considerations undoubtedly justify the supposition that intelligence may play a determining rôle in test reactions. Our immediate question, then, is concerned with its importance and its kind.

If intelligence is that type of ability which is measured by school grades and psychological tests, it does not function to an appreciable extent in motor accomplishment. Our evidence is again based upon coefficients of correlation. University grades, when correlated with scores in card-sorting, gave a coefficient of + .01; with Bogardus scores, − .17; with coördination scores, + .21. This last coefficient was the highest obtained between University grades and any motor test. The mental test scores showed a corresponding series of relationships. The fact should be noted, however, that the coefficients, although small, were positive; no inverse relationships were established.

These results by no means eliminate intelligence as a determining factor. But it is obvious that if we are to offer this factor as an explanation principle, our justification must be found in the definition given to the term. No common factor of any magnitude appears in the total group of measure-

ments recorded. The assumed intelligence factor, then, which functions in the reactions made to each test, seems to be peculiar to that test; it is specialized and particularized, and the term becomes, as a consequence, descriptive and not explanatory.

E. One of the most prominent features of the test behavior of the subjects was the manifestation of certain traits of character and temperament. These were clearly responses made to the test situations, even if they did include more than the specific motor adjustments elicited. It was thought that the particular set of such responses made by any one individual in the laboratory might be characteristic, in general, of that individual. If so, they might conceivably explain the specific reactions made to the tests, since the habitual general response would necessarily influence any specific response. Our task, then, was to ascertain the extent to which these general responses were characteristic of our subjects. The method employed has been described; but we may emphasize again the fact that the traits listed on the chart were the ones actually observed during the test periods.

The tabulations, based upon the scatter diagrams, would seem to indicate a positive relationship between proficiency in the tests and favorable ratings on the scale. But it should be noted that the method of tabulation employed tends to bring out relationships, since it groups the records into quartiles, while the coefficient of correlation, which takes into account all differences in rank, tends to minimize any actual gross correspondences which might actually exist. As regards positions above and below the medians, ratings and test scores do coincide to a noticeable degree. In order to make the results of the scatter diagram analysis comparable with the coefficients recorded, a number of relationships obtained by the latter method were measured by the scatter diagram method. This comparison shows a very slight tendency for character estimates to agree more closely than test measurements with other test measurements.

Evidently, the rating chart, as used in the present in-

vestigation, does not furnish the key for an explanation of motor ability. To reverse the statement, a test of motor skill cannot be used as a diagnostic measure of character and temperament. Perhaps the chief reason for this situation lies in the fact that we cannot assume with assurance the consistent manifestation of such traits in any individual. Poise, definitely manifested on one occasion, may break down on another. Our notes on behavior, as a matter of fact, contain instances of shiftings in emotional tone on the part of several subjects. And even more to the point is the question of the actual existence of the traits. It is possible that the poise manifested on a social occasion may be identical with the poise shown in the laboratory only as regards objective expression: the one may result from certain fixed habits, while the other may be the effect of an inhibitory process. The theory back of the rating chart itself may be responsible for its poor showing.

Yet a certain significance must be attached to these ratings. The median judgment on certain traits of an individual, furnished by five associates of that individual, even when recorded on a chart open to obvious criticisms, serves as a better indication of some forms of motor ability than other forms of ability, either motor or mental. We may reasonably expect then, that with a more adequate chart, better correlations would obtain. This chart would contain headings more closely descriptive of the reactions called for by the tests; and it would provide for variations in the amount of the trait under different conditions. It would be, in fact, a systematic description of such aspects of an individual's behavior as permit generalizations.

F. Although the conclusions resulting from the analysis of the data are seemingly negative in character, they may still contribute to a definite view of motor ability. For one thing, they enable us to question explanations of undue simplicity. In particular, they justify a certain amount of reserve as regards the term 'function.' No one would question the legitimate use of that term as applied to any one motor performance; but when it is assumed that generalized modes of

reaction exist, some doubt may reasonably be entertained. Phrenological categories are none the less dangerous when they are expressed in terms of objective behavior.

The net result of our investigation is the suggestion that even the simplest act of motor adjustment involves a play of factors not isolated through standard experimental precautions. One important group of such factors is undoubtedly that of transfer effects. It is useless to imagine that, while transfer may operate in card-sorting, it is completely eliminated in the Bogardus test, merely because the subject may acknowledge his familiarity with cards and his ignorance of the Bogardus apparatus. A second group of factors includes those ordinarily listed in connection with the learning process. The simplest motor test used probably elicited a genuine learning process; and this means that accidental success, for one thing, played an important rôle. That this factor operated may be inferred from the fact that successful and unsuccessful reactions were equally independent of the attempts of the subjects to analyze their problems. A third group of factors are those of the emotional and temperamental variety. Our reasons for not dismissing entirely the evidence furnished by the rating charts were given above. In addition, our notes hint strongly of the function of similar influences, not brought out by the charts. Inferiority complexes, shifting states of embarrassment and elation, stimulating and inhibiting ideas of unknown origin and character, were only too evident. The psychology of motor accomplishment awaits the formulation of a psychology of motivation.

References

1. Brown, Warner. *Habit Interference in Sorting Cards.* Univ. of Calif. Pub., Vol. 1, No. 4.
2. Thorndike, E. L. *Mental and Social Measurements.* Second Edition.
3. Thurstone, L. L. Mental Tests for College Entrance. *Jr. of Ed. Psychol.*, Vol. X., No. 3.
4. Wissler, Clark. The Correlations of Mental and Physical Tests. *Psychol. Monog.*, Vol. III., No. 6.

COMPARATIVE SIMPLE REACTIONS TO LIGHT AND SOUND

From the Psychological Laboratory of McLean Hospital[1]

F. L. WELLS, Ph.D., In Charge
C. M. KELLEY, M.D., Associate
GARDNER MURPHY, A.M., Assistant

1. It is well known that under ordinary laboratory conditions the simple reaction time to sound stimuli is shorter than to light stimuli, though reactions to minimal stimuli seem to be equal (Wundt). Richet's (1) calculations give a mean of 149 for sound and 197 for light stimuli. Recently Dockeray and Isaacs report a similar finding, 155 to 197 sigma, based on series of 25 observations each with candidates for flying service. It also seems that reaction times become shorter as the magnitude of the stimulus is increased. In this case any relationship between sound and light stimulus depends on the relative magnitude of the sound and light stimuli. In Froeberg's (2) results, different relationships thus appear according to the magnitudes of the stimuli selected for comparison. In his subject *W*, who is the present writer, this sound-light ratio (per cent. which the light time is of the sound time) ranges from 90 to 158, with different relative magnitudes of sound and light stimulus. Theoretical equality in magnitude might be established by taking a sound and light showing equal number of just noticeable differences above the threshold, though the difficulties seem all but insuperable. For example, the sound-light ratios in Woodrow's (3) subjects *Ht*, *Vs* and *Sz*, for 'weak' intensities of stimuli, are 111, 139 and 128 respectively, their ratios for 'medium' sound and 'bright' light are 136, 140 and 136 respectively.

2. Apparatus for giving standard stimuli was not available for the present experiments. The sound stimulus was the

[1] As of August 1, 1920.

click of opening an ordinary 20-ohm relay. It is clearly audible, but a weaker stimulus than ordinarily used. Circuit was made through the chronoscope at the moment of impact. The subject is 7 feet distant from the source of the stimulus. The incidence of the stimulus is thus about 6 sigma later than the impact, and this quantity is deducted from the readings.

3. The light stimulus is the darkening of a tungsten bulb by interrupting the current through it. The same interruption establishes the chronoscope circuit, approximately 6 sigma after the interruption. How long after this interruption the dimming of the light becomes effective as a stimulus is uncertain (4). The incidence of stimulus is treated as synchronous with the beginning of timing.

4. The material consists of 33 experiments of 216 reactions each. There are 13 subjects. Eleven subjects, one woman, had only 216 reactions each. The two other subjects, *K* and *W*, had 2,376 reactions each. The term 'A-group' is used to denote the single experiments made with the 11 subjects, combined with the first experiments with subjects *K* and *W*. The single experiment of 216 reactions consists of 108 reactions to sound and 108 reactions to light. Throughout the experiments, small and irregular numbers of sound reactions alternate with small and irregular numbers of light reactions. This was so that practise, or other intercurrent changes might not have undue differential effect on either class of reactions. The subject is forewarned when to expect a group of sound or a group of light stimuli. When 54 reactions of each kind have been taken, there is a pause of perhaps 90 seconds, after which the remaining half of the experiment is completed. The 216 reactions occupy about 40 minutes; the method does not lend itself to taking reactions in very rapid succession. The experimenter's time is economized by distributing the reactions as they are read.

5. The reaction times were measured with a Leeds and Northrup Type H D'Arsonval galvanometer arranged as a chronoscope, in a manner derived from suggestions by Klopsteg (5). This method has been very satisfactory. The

simple reaction observations to date are upwards of 25,000. Of these 7,128 are directly concerned with the present results. The remainder are mostly with pathological cases.

6. The reaction movement was that of releasing, with the right hand, an ordinary telegraph key. In the light reactions of the second experiments with *K* and *W*, correction is made for a constant error approximating plus 107 sigma, occasioned through a tampering with the set-up. The effect on the averages is negligible.

7. The score of each experimental unit (a series of 27 reactions) is the median. Aside from its economy of time, its use is almost compulsory for pathological material and thus for data to be compared with it. The present concern is with the relations of sound time and light time. This is expressed, as in par. 1, by a ratio, the per cent. which the light time is of the sound time. General average of sound time scores for the A-group is 204 sigma, a little less than the average for Froeberg's 16 millimeter drop. The corresponding light time is 237 sigma, some 21 sigma more than for Froeberg's weakest light stimulus. Their ratio as above (termed the 'sound-light ratio') is thus 116 as against 132 for Richet's figures, where the stimuli were apparently farther apart. In the stimuli used by Dunlap and G. R. Wells (6) still greater disparity appears.

8. Individual differences may be said to exist where a number of records taken under similar conditions on different individuals clearly differ more than a similar number of records made with a single individual. The present material compares the results of single experiments with 13 individuals, the A-group, with 11 experiments on each of two individuals, *K* and *W*. The sound-light ratio, *i.e.*, the per cent. which the light time is of the sound time, is as follows in the three groups of data:

	A-group	*K*	*W*
Median	115	134	146
Average	116	133	145
M.V.	7.8	4.5	3.2

9. If the mean diurnal variations in *K* and *W* are representative, the relation of sound time to light time is a function of individual difference, less distinct than the tapping test, and of about the same order of distinctness as the æsthetic and color judgments of a previous study (7). Further consideration is invited by the extreme positions of the ratios in *K* and *W*.

10. One factor to think of is instrumental. The experiments with *K* and *W* were made after those with the other subjects. Alterations were meanwhile made in the set-up. The question resolves into whether the light stimulus could so have been delayed beyond the beginning of timing. Except for the accidental condition mentioned in par. 6, the set-up seems guarded against this. Control calculations from ten records on pathological subjects made with the initial set-up, and ten other pathological records made with the *K* and *W* set-up (4,230 observations), show medians the same in each, 112; av. 112, m.v. 6.5, av. 118, m.v. 11.3 respectively, one outlying case in second group. One patient is in both groups. His ratio in the first is 106, in the second, 103.

11. Some explanation might be sought in the reaction history of the subjects. Those of the A-group, excepting *K* and *W* (all but one nurses) were without special practise so far as known. Previous to 1907 *W* had much practise to both sound and light reactions. *K*'s previous practise is negligible, as is also the practise effect in the experiments themselves. The situation seems equally unsuited to interpretation in terms of differential practise. In imagery type, *K* is a visualizer, *W* auditory-motor.

12. If *K* and *W* present idiosyncrasies in this respect, it may be related to the fact that both have always shown rather short reaction times to sound. The physiology of sound and light stimuli, and the complication of the latter with inertia effects, suggest that there may be a relatively lower 'physiological limit' for the sound-reaction. This would account for larger sound-light ratios in those with short reaction times. In the A-group, correlation of sound time with the sound-

light ratio is $\rho = -.52$. Those with quicker time to sound tend towards *relatively* slower time to light.

13. It follows that smaller individual differences should be expected in light than in sound reactions. Between members of the A-group the light variation is about 75 per cent. of that for sound. In Froeberg's work, relative variation of single series also, is notably more for sound than for light. This is the case with *K* and *W* in the present material, but in the A-group there is no difference of this kind.

14. The electrical conditions of the experiment necessitate a get-ready signal. Mechanical control of the prestimulus interval presented problems it was not practicable to solve for these conditions. By listening to a stopwatch, the operator kept the prestimulus interval in half the cases, irregularly interspersed, as close to one second, in half the cases as close to three seconds, as he could. It is possible that this necessity is a virtue (8). Special measurements indicated the following properties for the prestimulus intervals:

Length in seconds of the prestimulus intervals with the two experimenters.

	"One Second"		"Three Seconds"	
	Av.	M.V.	Av.	M.V.
Experimenter *K*	1.21	.09	3.10	.18
" *W*	1.20	.11	3.22	.14

These are the intervals which in further allusions will be known as the 'one-second' and 'three-second' intervals respectively. *W* was the experimenter in all the present observations except those in which he was subject.

These prestimulus intervals make no difference in the averages of the A-group, or with *W*. The sound-light ratios are there practically equal, with either prestimulus interval. In the case of *K* it makes a distinct difference, and so, that the light stimulus is relatively favored by the longer prestimulus interval. His figures are as follows:

Sound-light Ratios in Subject *K*	Av.	M. V.
One second prestimulus, first half of experiments	142	8.6
One second prestimulus, second half of experiments	136	7.1
Three seconds prestimulus, first half of experiments	131	5.8
Three seconds prestimulus, second half of experiments	126	5.7

15. To sum up: The relation between reaction-time to sound and light is dependent on the magnitude of the stimuli compared. An absolute sound-light ratio might be formulated only on the basis of stimuli having the same sensation value for different subjects. Comparison of the reaction times of a relatively fixed sound and light stimulus, shows individual differences in their respective values for the sensation-reaction process. Subjects with relatively short sound reaction times do not have light reaction times decreased in the same degree. Correspondingly, there is somewhat more individual difference in reaction time to sound than to light. Idiosyncrasy in the sound-light ratio is observed, in respect to special favoring of sound, and in respect to the effect of different prestimulus intervals.

REFERENCES

1. Richet. Dictionnaire de Physiologie, 1898, **3**, 19.
2. Froeberg, S. The Relation between the Magnitude of the Stimulus and the Time of Reaction. *Arch. of Psychol.*, 1907, No. 8, pp. 38.
3. Woodrow, H. Reactions to the Cessation of Stimuli and their Nervous Mechanism. *Psychol. Rev.*, 1915, **22**, 423–452, esp. p. 432.
4. Johnson, H. M. The Influence of the Distribution of Brightnesses over the Visual Field on the Time Required for Discrimination Responses to Visual Stimuli. *Psychobiology*, 1918, **1**, 459–494, esp. p. 468.
5. Klopsteg, P. E. A New Chronoscope and Fall Apparatus. Journ. Exp. Psychol., 1917, **2**, 253–263.
6. Dunlap, K., and Wells, G. R. Experiments with Reactions to Visual and Auditory Stimuli. *Psychol. Rev.*, 1910, **17**, 319–336.
7. Wells, F. L. On the Variability of Individual Judgment. Essays in Honor of William James, 1908, 511–549.
8. Woodworth, R. S. Elements of Physiological Psychology. 1911, p. 481. Quoted by Todd, J. W., Reactions to Multiple Stimuli, *Arch. of Psychol.*, **1912, 25**, p. 11.

AN EXPERIMENTAL STUDY OF THE "STIMULUS ERROR"

BY SAMUEL W. FERNBERGER

University of Pennsylvania

Titchener has clearly made the distinction between the stimulus attitude and the process attitude in the formation of sensory judgments.[1] In the case of the process attitude, the observer is instructed to judge process only and, indeed in most cases, only one attributive aspect of a given process. This, Titchener insists, is the only proper psychological attitude for experimental work and is obtained only with observers, trained in introspection, who have been given very carefully worded instructions so that the process to be reported is absolutely unequivocal. Under the stimulus attitude, on the other hand, the observer judges not process, but the stimuli themselves. This, Titchener insists, is therefore not psychological, but is rather on the meaning level and gives judgments of an equivocal sort. Hence he terms it, not the stimulus attitude, but the 'stimulus error.'

The equivocality of the judgments under the stimulus attitude is well shown for the case of the two-point cutaneous limen by Titchener in a later paper.[2] He here points out that in this sort of an experiment there exist several perceptive forms between exact 'one-ness' and clear 'two-ness.' Under the process attitude all of these intermediate forms are reported, while under the stimulus attitude, all of these have the meaning of 'more-than-one-stimulus,' and are therefore

[1] Titchener, E. B., 'Text-Book of Psychology,' 1910, pp. 202f., 218.

[2] Titchener, E. B., 'On Ethnological Tests of Sensation and Perception with Special Reference to Tests of Color Vision and Tactile Discrimination Described in the Reports of the Cambridge Anthropological Expedition to Torres Straits, *Proc. Amer. Philos. Soc.*, 1916, **55**, 204–236. For a discussion of the equating of stimulus attitude with equivocal determination, *cf.*, Boring, E. G., The Control of Attitude in Psychophysical Experiments,' *Psychol. Rev.*, 1920, **27**, 440–452.

reported as 'two.' Hence, under the stimulus attitude, the two-point limen will be smaller and the observer will appear to have a greater sensitivity.

Curiously enough the discussion on these points has been largely epistomological and very little experimental work bears directly on the distinction. Titchener bases his discussion of the Torres Straits results on the experiments of Gates[1] and of others who determined the intermediate forms. These experiments were performed from another viewpoint and in more detailed fashion by de Laski and later by Friedline.[2] These investigators, both working with the two-point limen, take into account the intermediate perceptual forms between 'one-ness' and 'two-ness' and emphasize the importance of such a consideration in working out the statistical limen for duality.

The use of lifted weight stimuli presents a situation in which the distinction between the two attitudes should be important. All of the earlier work in this field has been performed under the stimulus attitude—the observers have been instructed to judge the differences between the weights themselves. In the case of lifted weights, it has been shown that the processes involved in the formation of the judgment are exceedingly complex and that these processes actually vary from individual to individual observer and for the same observer at different stages of practice.[3] It has been found that pressure sensations on the tips of the fingers, kinæsthetic sensations localized in the hand, wrist and forearm might all be involved. In the early stages of practice, the pressure criteria seem to be most used but, after considerable progressive practice, most observers tend to form their judgments on the basis of kinæsthetic criteria. Working under the stimulus attitude, the observers seem to use a cri-

[1] Gates, E. J., 'The Determination of the Limens of Single and Dual Impression by the Method of Constant Stimuli.' *Amer. Jour. of Psychol.*, 1915, **26**, 152–157.

[2] de Laski, E., 'On Perceptive Forms Below the Level of the Two-Point Limen,' *Amer. Jour. of Psychol.*, 1916, **27**, 569–571. Friedline, C. L., 'The Discrimination of Cutaneous Patterns Below the Two-Point Limen,' *Amer. Jour. of Psychol.*, 1918, **29**, 400–419. A description of the earlier experiments will be found in Friedline's paper.

[3] Fernberger, S. W., 'An Introspective Analysis of the Process of Comparing.' *Psychol. Mono.*, 1919, No. 117, pp. 161.

terion of higher order and one which is on the perceptual level—that of the second weight going up 'higher' or 'faster,' 'lower' or 'slower' than the first weight—a perceptual criterion, probably compounded of temporal, intensity and extensity (of localization) attributive kinæsthetic components.

Here then would seem to be an excellent field in which to discover whether the judging under stimulus attitude or the judging of some particular attributive process will have any effect upon the statistical limens. This problem has been recently attacked, from one point of view, by Friedländer.[1] He employed lifted weights as his stimuli and, by means of instructions, set up the following different attitudes in his observers: (1) An attitude of judging 'pressure sensations on the inner side of the fingers,' (2) an attitude of judging strain or 'Kraft' sensations, (3) a general attitude of judging the fused perception representing the stimulus (*cf.* pp. 134f.). The lifting was performed for all three series with and without gloves in an effort to ascertain if this change in the objective conditions had any influence on the frequency of the different judgments for the different series. Friedländer finds that, in the pressure attitude series, the presence of a glove on the lifting hand has a great influence on the judgments, while it has little influence in the other two series. He also compares the sensitivity under the objective and subjective attitude and concludes (p. 191) that the sensitivity under the objective attitude is somewhat greater than under the subjective.

Friedländer has shown, then, that it is possible to have the observer adopt either an objective attitude or an attitude of judging only one of several elements in a complex, and, that the adoption of one or other of these attitudes may effect the threshold values. It would appear, however, that his instructions for the subjective attitudes (especially the kinæsthetic) were still too indefinite and equivocal and were therefore capable of further refinement. Also his statistical conclusion regarding the effects of the attitudes upon the

[1] Friedländer, H., 'Die Wahrnehmung der Schwere,' *Zeit. f. Psychol.*, 1920, 83, 129–210.

measure of sensitivity are not conclusive, inasmuch as he reports in this connection the results of only one observer whose reactions were made upon comparison stimuli with too wide an interval (we believe) and who gave too few judgments for them to be of great statistical value.

It seemed then worth while to repeat these experiments with more exact and less equivocal instructions; to use more observers and to use observers at different stages of practice; and to obtain more judgments on a more appropiately arranged series of stimuli.

The stimuli consisted of a standard weight of 100 grams and comparison weights of 88, 92, 96, 100 and 104 grams. Hard rubber stimuli were used so as to eliminate the temperature sensations.[1] These weights were presented in the first time order: *i.e.*, with the standard stimulus always presented first, and with the space error eliminated. A double series of weights were presented in a carefully arranged order so as to eliminate, as far as possible, the effects of the order of the series.[2] The manner of lifting has been elsewhere described in detail.[3] The only difference between the manner of lifting in the earlier experiment and in the present one is that, in the former, the lifting was continuous and the time between the pairs was kept constant while, in the present experiment, the observer was allowed to stop between each pair if he so desired.

Three observers took part in the experiment. Mr. Carroll C. Pratt, A.M., assistant in psychology (Ob. *P*) had a great deal of practice in experimentation of this sort and under these objective conditions, having taken part in a similar experiment last year. Miss Marjory Bates, A.M. (Ob. *B*), graduate student in psychology, had a good deal of introspective training and had some little practice in lifted weight experimentation in the laboratory course last year. Mr.

[1] Fernberger, S. W., 'A New Form of Stimuli for Lifted Weight Experiments,' *Amer. Jour. of Psychol.*, 1920, 31, 147–151.

[2] Fernberger, S. W., 'Interdependence of Judgments Within the Series for the Method of Constant Stimuli,' Jour. of Exper. Psychol., 1920, 3, 126–150.

[3] Fernberger, S. W., 'On the Relation of the Methods of Just Perceptible Differences and Constant Stimuli,' *Psychol. Mono.*, 1913, No. 61, 6–11.

Henry M. Halverson, A.M. (Ob. *H*), graduate student in psychology, was totally unpracticed in this sort of experimentation but had a certain amount of introspective training. The experiments were performed in the laboratory of experimental psychology, Clark University, during the autumn of 1920. All of the observers were given sufficient preliminary practice so that the hand movements became automatic.

Three series of experiments were obtained from each of the three subjects. In the first series (Series *A*) the following instructions were given to the observers.

> Certain stimuli will be presented to you which will be lifted—in so far as the rhythm and the manner of lifting is concerned—in accordance with your former instructions.
>
> You will report only on the intensity of the pressure sensations on the tips of the fingers. Attend to these pressure sensations only. Three categories of judgment are permitted—'greater,'' less' and 'equal.' A greater or a less judgment indicates that the intensity of the pressure sensations on the tips of the fingers aroused by the second stimulus, was greater or less respectively than the intensity of the pressure sensations aroused by the first stimulus of the same pair. An equal judgment indicates that the intensities of the pressure sensations aroused by the two members of a pair were subjectively equal.
>
> If you fail to live up to these instructions, either with regard to the categories of judgment or with regard to the formation of the judgment on the basis of the intensity of the pressure sensations only, you will not give a judgment but will indicate the fact by reporting the word 'failure.'
>
> Intensity of pressure sensations in the above is to be taken in the sense of mental process only.

The second series of experiments (Series *B*) was made under instructions which imposed the judging of the intensity of the kinæsthetic sensations localized in the wrist. These instructions were exactly similar to those for Series *A*, except for the insertion of the words 'intensity of kinæsthetic sensations localized in the wrist' for the words 'intensity of pressure sensations on the tips of the fingers.'

The third series of experiments (Series *C*) was made under instructions which were planned to arouse the stimulus attitude in the observers and which are here given in full.

> Certain stimuli will be presented to you which will be lifted—in so far as the rhythm and the manner of lifting is concerned—in accordance with your former instructions.
>
> You are to take the stimulus attitude and are to judge the weights themselves. Three categories of judgment are permitted—'greater,' 'less' and 'equal.' A greater

or a less judgment indicates that the second weight was greater or less respectively than the first weight of the same pair. An equal judgment indicates that the two stimuli of a pair were of equal weight.

If you fail to live up to these instructions for any reason you will not give a judgment but will indicate the fact by reporting the word 'failure.'

By means of these different instructions it was hoped that the observers would make an attentional abstraction from the mass of processes involved and, in Series *A* and *B*, judge only the intensity of the pressure sensations on the finger

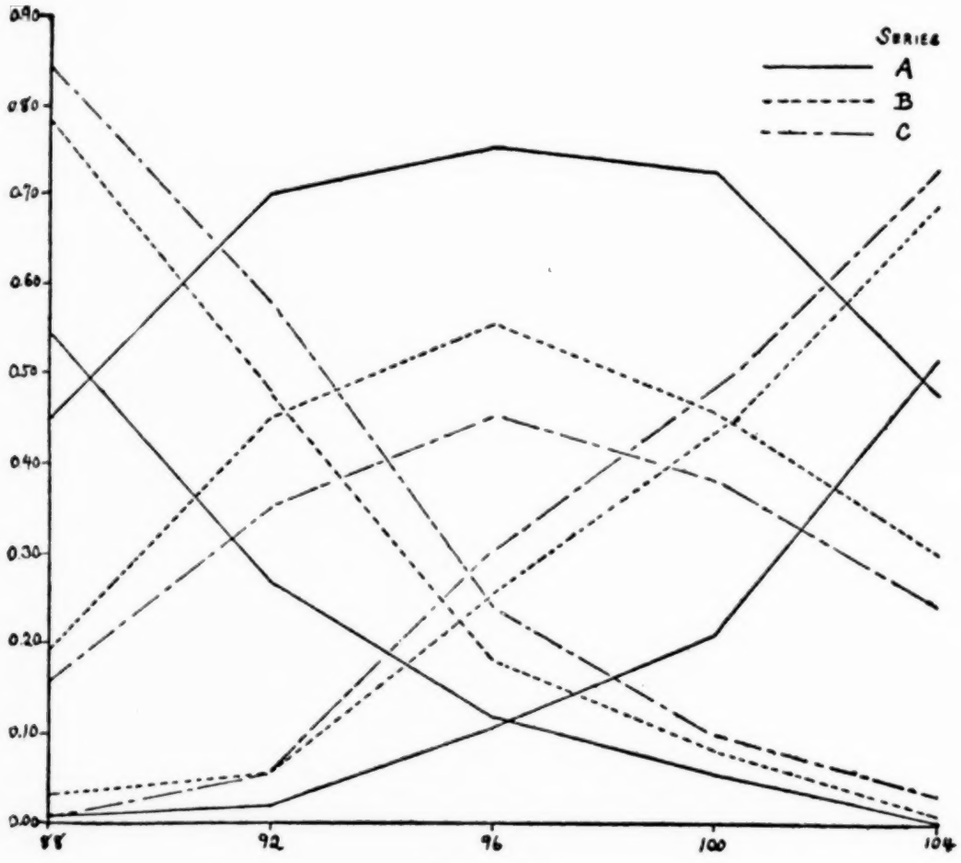

Chart I.

Observer P.

tips and the intensity of the kinæsthetic sensations in the wrist respectively. The statistical values for these series could be then compared with one another and with those of Series *C*, taken under stimulus attitude instructions. Introspective characterizations were taken for every series so that it could be assured that the observers were making the

proper attentional abstractions and were forming the judgments on the basis of the processes desired. An examination of these introspective characterizations shows that the instrucions were properly fulfiled.

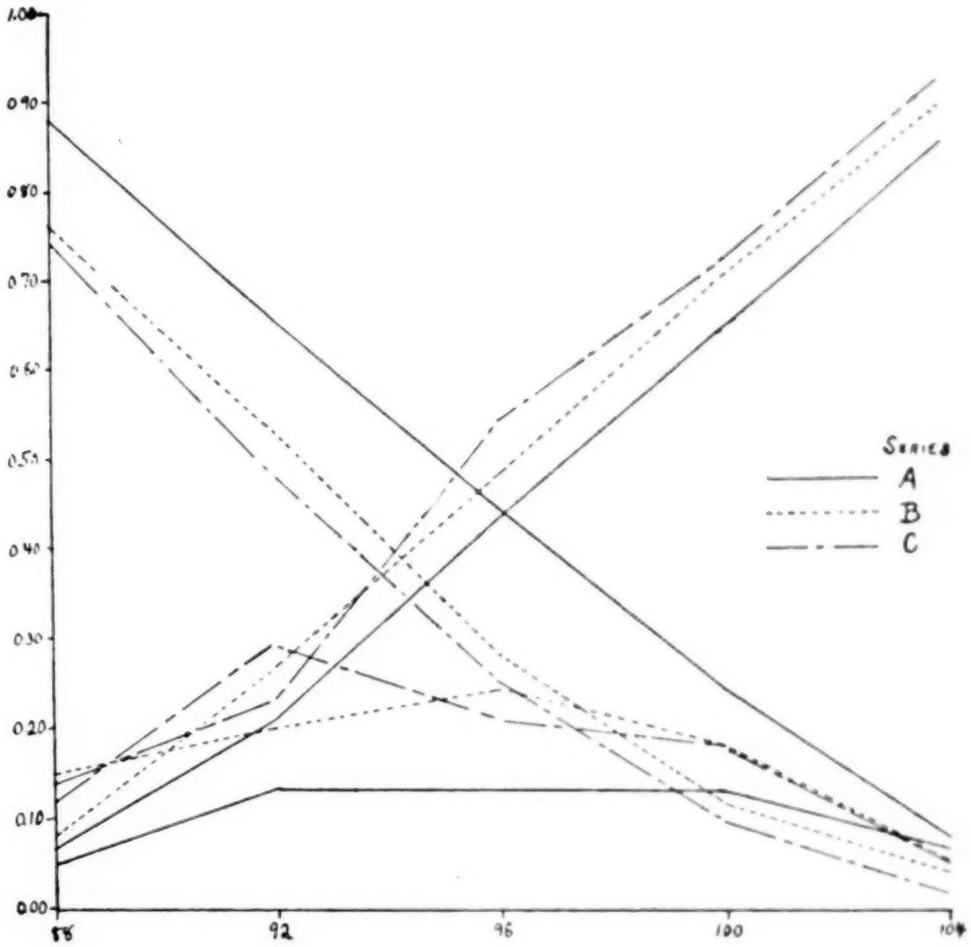

CHART II.
Observer B.

In each series 250 judgments were taken from each observer on each of the five stimulus pairs. The calculations were made in accordance with the $\Phi(\gamma)$ hypothesis by the methods developed by Urban.[1]

The observed relative frequencies for the different series are plotted in the form of curves in Charts I.–III. Each chart contains the curves for all three series for one subject.

[1] Urban, F. M., Hilfstabellen für die Konstanzmethode. *Arch. f. d. ges. Psychol.*, 1914, 21, 337–340.

The curves for Observer *P* are given in Chart I.; those for Observer *B* in Chart II.; and those for Observer *H* in Chart III. The charts are similarly constructed. The values of the comparison stimuli are laid off along the abscissa and the observed relative frequencies are erected as ordinates. The curves for Series *A* (intensity of pressure sensations on the finger tips) are represented by the solid lines. The curves for Series *B*. (intensity of kinæsthetic sensations in the wrist) are represented by the short dash lines. The curves for Series *C* (stimulus attitude) are represented by the alternating short and long dash lines.

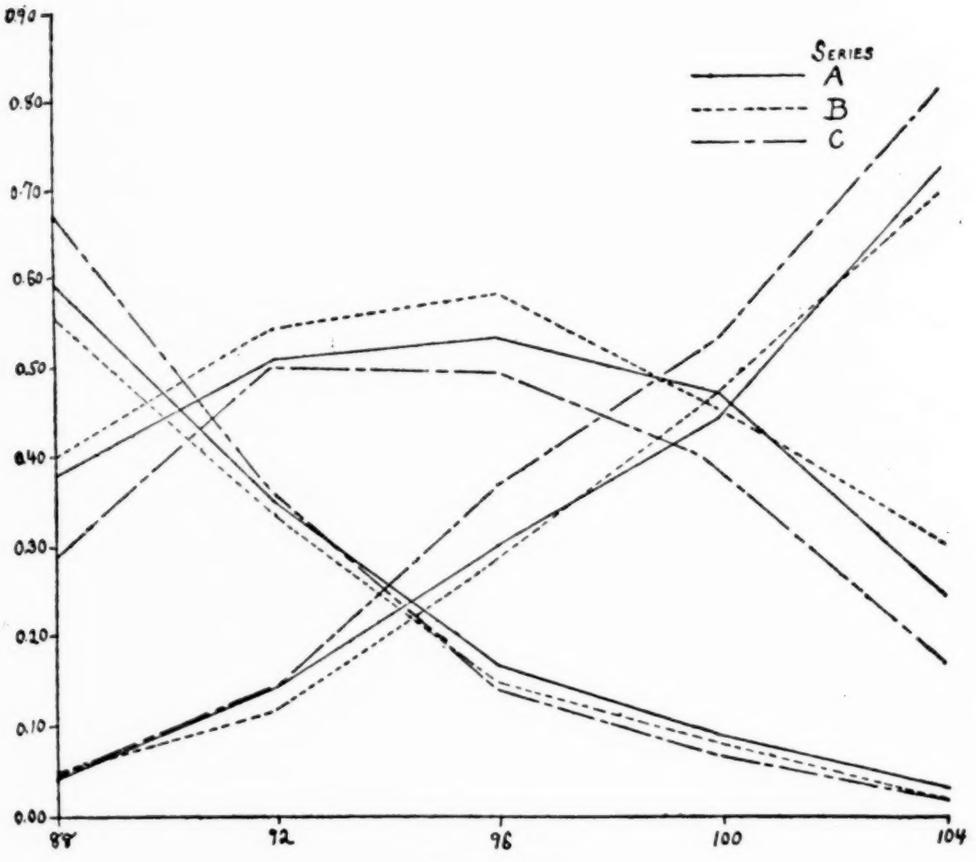

Chart III.

Observer H.

An examination of these charts indicates that all of the curves for every series are of a form which, from inspection, is relatively similar to the form of the Φ (γ) hypothesis. In no case is there a single inversion. This is evidence that

each of the instructions given the observers are proper instructions under which to acquire data which may be properly treated, statistically, by the method of constant stimuli. The introspections from the observers gave additional evidence that they were able to make the attentional abstraction required by the instructions and that they were able to form judgments on the basis of them.

It will be observed, however, that these curves vary in form and size for the different series for the same observer—in some cases to a very considerable extent (cf. the equality curves for Ob. *P*, Chart I.). These differences can be best discussed by considering the magnitudes of the constants of the psychometric functions, of the thresholds, the intervals of uncertainty and the points of subjective equality for the different series for each observer.

TABLE I

OBSERVER *P*

Series	h_1	h_2	S_1	S_2	Interval of Uncertainty	Point of Subjective Equality
A.....	0.102	0.103	88.24	104.49	16.25	96.06
B.....	0.131	0.108	91.83	100.68	8.85	95.84
C.....	0.131	0.117	93.01	99.85	6.85	96.26

TABLE II

OBSERVER *B*

Series	h_1	h_2	S_1	S_2	Interval of Uncertainty	Point of Subjective Equality
A.....	0.111	0.108	95.08	97.52	2.45	96.37
B.....	0.112	0.109	92.57	96.33	3.76	94.43
C.....	0.112	0.111	91.89	95.80	3.91	93.83

TABLE III

OBSERVER *H*

Series	h_1	h_2	S_1	S_2	Interval of Uncertainty	Point of Subjective Equality
A.....	0.095	0.097	89.29	100.27	10.98	94.85
B.....	0.093	0.097	88.83	100.49	11.66	94.75
C.....	0.118	0.112	90.26	98.76	8.50	94.35

These values are given in Tables I.–III. Each table is similarly constructed and contains the results of the three series for one subject. In the first column are given the letters of the series. The second and third columns contain the values of the coefficients of precision of the less judgments (h_1) and of the greater judgments (h_2) respectively. The fourth and fifth columns contain the values of the lower (S_1) and the upper (S_2) thresholds respectively. The sixth column contains the values for the interval of uncertainty (S_2–S_1). The last column contains the values for the point of subjective equality

$$\left(\frac{c_1 + c_2}{h_1 + h_2}\right)$$

and indicates the point on the abscissa corresponding to the point of the intersection of the curves of the psychometric functions of the less and greater judgments. In the rows are found the average values for all of the columns for Series *A*, *B* and *C* respectively.

The results for Observer *P* (Table I.) will be first considered. The values for the coefficients of precision for the greater judgments (h_2) are very similar for all three series. The coefficients of precision for the less judgments (h_1) for Series *B* and *C* are identical; while both are considerably larger than that for Series *A*. The thresholds in the direction of decrease (S_1) vary considerably—the lowest being for Series *A*, the highest for Series *C*, and that for Series *B* being intermediate. The threshold in the direction of increase (S_2) for Series *A* is considerably larger than those for either of the other two series; while that for Series *B* is slightly larger than the value for Series *C*. The values of the interval of uncertainty for the different series in *P*'s results are quite different. The interval for the pressure series (*A*) is more than twice as large as that for Series *C* and nearly twice as large as that for Series *B*. Also the value for Series *B* is nearly one third as large again as that for Series *C*—the actual difference being 2 grams. The values, from Observer *P*'s results, for the points of subjective equality are very similar in all three cases—the greatest difference (*B* and *C*) being 0.42 grams.

In the case of the results of Observer *B* (Table II.), the following may be noted. The coefficients of precision for both the psychometric functions of the less (h_1) and greater (h_2) judgments are, for all three series practically identical. The value of the lower threshold (S_1) for Series *A* is considerably higher than those for Series *B* and *C*; while the *B* and *C* values are practically identical. The values for the upper thresholds (S_2) are very similar but that for Series *A* is higher than those for Series *B* and *C*; while that for Series *B* is slightly higher than the value for Series *C*. The values of the interval of uncertainty for Series *A* is smaller than those for either *B* or *C*. The values for Series *B* and *C* are very similar. Similar relations hold with regard to the points of subjective equality for this observer. The value for Series *A* is considerably higher than those for the other series. The differences between the points of subjective equality of Series *B* and *C* is relatively small (0.60).

For Observer *H* (Table III.) the coefficients of precision for the psychometric functions of the less (h_1) and the greater (h_2) judgments are very similar for all three series. The values of the lower thresholds (S_1) are relatively similar for all three instructions—that for Series *B* being the lowest and being very similar to that for Series *A*. The value for Series *C* is somewhat higher than either of the others. The value of the upper threshold (S_2) for Series *C* is considerably lower than those for either Series *A* or *B*, while the values for *A* and *B* are almost identical. The value of the interval of uncertainty for Series *C* is considerably smaller than those for the other two series. The values of the interval of uncertainty for Series *B* and *C* are not very different. The points of subjective equality for all three series are quite similar, the greatest difference being only 0.50 grams (*A* and *C*).

This then appears to be a curious situation when one considers the effects of the various instructions for the different observers for the values of the intervals of uncertainty and for the points of subjective equality. These differences cannot be understood, we believe, without also considering the stage of progressive practice in which we find each of

the subjects. It has been shown elsewhere[1] that the processes underlying the formation of judgments of small differences in weight taken under the stimulus attitude vary for different stages of practice. For the unpracticed subject, the judgment is made largely on the basis of pressure sensations on the tips of the fingers, while as practice continues, these are gradually lost until there is a complete shift to kinæsthetic criteria.

In the light of these findings only would it seem possible to interpret the present results. Observer *H* was the totally unpracticed observer. For him the series taken under the stimulus attitude gives the smallest interval of uncertainty. The interval of uncertainty for the kinæsthetic instructions is greater than that for pressure. This result might have been expected at that stage of practice inasmuch as pressure seems to be the 'natural' early criterion. Observer *P* was the highly practiced subject and for him the relations are reversed. Again the stimulus attitude series gives the smallest interval of uncertainty. But the introspections of this observer show that, under the stimulus attitude (Series *C*), he judged largely in terms of kinæsthesis and the interval for his kinæsthetic instructions (Series *B*) much more nearly approximates the value for the stimulus attitude (Series *C*) than they do the pressure instructions (Series *A*). The values for Observer *B* are somewhat equivocal. She was also relatively unpracticed and, as in the case of the other unpracticed observer *H*, the interval of uncertainty for the pressure series (*A*) is smaller than for the kinæsthetic series (*B*). Thus far all is as one might expect it to be. But then the value under the stimulus attitude (*C*) is greater than those for either of the other series. This subject states, in her introspection, that the processes for Series *C* were complex—involving tactual and widespread kinæsthetic experiences and a great deal of visual imagery. She also states that the judgments in this series were given with a much lower degree of subjective assurance, on the whole, than in either of the other series. It is possible that this observer was just in the transitional

[1] Fernberger, S. W., 'An Introspective Analysis of the Process of Comparing,' *Psychol. Mono.*, 1919, No. 117, 136–144.

stage between early practice and late practice. Certainly her results show characteristics of early practice inasmuch as the interval of uncertainty for the kinæsthetic series is larger than that for the pressure series. If she were in the transitional stage, it might well be that in the stimulus attitude series, there was a conflict between the partially disregarded pressure criteria, of the early practice, with the newly acquired but not yet perfected kinæsthetic criteria. This would lead to a certain amount of equivocality in the judgments which would show in a greater number of equality judgments—and hence the larger interval of uncertainty—and also in the lack of subjective assurance reported in the introspections.

The values for the points of subjective equality are very constant for Observer *P*, who was the practiced observer. The values for Observers *B* and *H* show a regular tendency to decrease in the order in which the series were taken. This decrease in the size of the point of subjective equality is one of the normal phenomena of progressive practice as has been pointed out elsewhere.[1] The effects are very slight for Observer *H* and very considerable for Observer *B*. Hence it would seem that the results of these two observers, with respect to the value of the points of subjective equality cannot throw light on the present problem, inasmuch as the effects of progressive practice might have cancelled any effects of the instructions. From the results of Observer *P*, it would appear as if the instructions employed had little if any effect upon the value of the point of subjective equality.

Conclusions

1. It is evident that a trained observer is able, by an attentional abstraction and isolation, to judge intensity differences of a single modality from a complex in which several modalities of sensation are present simultaneously.

2. In the case of lifted weight stimuli, the following instructions are capable of fulfilment; namely, judging the intensity of the pressure sensations on the tips of the fingers;

[1] Fernberger, S. W., 'The Effects of Practice in Its Initial Stages in Lifted Weight Experiments and Its Bearing upon Anthropometric Measurements,' *Amer. Jour. of Psychol.*, 1916, 27, 261–272.

the intensity of the kinæsthetic sensations localized in the wrist; and judging, by assuming a stimulus attitude, the relative intensity of the weights themselves. This latter is distinctly a complex perceptual process on the 'meaning' level.

3. All three of these instructions give curves of the psychometric functions which appear to be somewhat similar in form to the $\Phi(\gamma)$ hypothesis.

4. The pressure and kinæsthetic instructions give rise to strictly psychological results. Whether or not the stimulus attitude instructions are strictly psychological or belong to an experimental logic depends on one's definition of psychology. Certainly all of the classical experiments in the field of lifted weights have been performed under this attitude.

5. The coefficients of precision for the psychometric functions of both the less and greater judgments are very similar for all three instructions.

6. The size of the intervals of uncertainty vary for the different instructions. For the unpracticed subjects, the interval for the pressure sensations is smaller than for the kinæsthetic sensations. For the practiced subject, the interval for the kinæsthetic sensations is smaller than that for the pressure sensations. There seems to be a tendency for the interval obtained under the stimulus attitude to be smaller than either those for the pressure or kinæsthetic sensations alone. The stimulus attitude, however, may lead to an equivocal sort of judgment which results in a low degree of subjective assurance; which in turn results in a tendency to increase the number of equality judgments and thus leads to a larger interval of uncertainty.

7. The point of subjective equality does not seem to be effected by the different instructions. The time error is evident for pressure, kinæsthesis and the stimulus attitude. Progressive practice, evidenced by its effect in lowering the point of subjective equality, is evident throughout all the series for the unpracticed observers.

8. This experiment seems to show the futility of properly interpreting the statistical results of a purely behavioristic study without the control of introspective report.